OLUME 102 PART C NUMBER 1 MARCH 1955

The Proceedings

OF

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART C

MONOGRAPHS Nos. 100–116

SAVOY PLACE · LONDON W.C.2

Price Eight Shillings and Sixpence

The Institution of Electrical Engineers

FOUNDED 1871

INCORPORATED BY ROYAL CHARTER 1921

PATRON: HER MAJESTY THE QUEEN

COUNCIL 1954–1955

President

J. ECCLES, C.B.E., B.Sc.

Past-Presidents

SIR JAMES SWINBURNE, BART., F.R.S.
W. H. ECCLES, D.Sc., F.R.S.
THE RT. HON. THE EARL OF MOUNT EDGCUMBE, T.D.
J. M. DONALDSON, M.C.
PROFESSOR E. W. MARCHANT, D.Sc.
P. V. HUNTER, C.B.E.
H. T. YOUNG.
SIR GEORGE LEE, O.B.E., M.C.
SIR ARTHUR P. M. FLEMING, C.B.E., D.Eng., LL.D.
J. R. BEARD, C.B.E., M.Sc.
SIR NOEL ASHBRIDGE, B.Sc.(Eng.).
COLONEL SIR A. STANLEY ANGWIN, K.B.E., D.S.O., M.C., T.D., D.Sc.(Eng.).
SIR HARRY RAILING, D.Eng.
P. DUNSHEATH, C.B.E., M.A., D.Sc.(Eng.).
SIR VINCENT Z. DE FERRANTI, M.C.
T. G. N. HALDANE, M.A.
PROFESSOR E. B. MOULLIN, M.A., Sc.D.
SIR ARCHIBALD J. GILL, B.Sc.(Eng.).
SIR JOHN HACKING.
COLONEL B. H. LEESON, C.B.E., T.D.
H. BISHOP, C.B.E., B.Sc.(Eng.).

Vice-Presidents

T. E. GOLDUP, C.B.E.
S. E. GOODALL, M.Sc.(Eng.).
WILLIS JACKSON, D.Sc., D.Phil., F.R.S.
SIR GEORGE H. NELSON.
SIR W. GORDON RADLEY, C.B.E., Ph.D.(Eng.).

Honorary Treasurer

H. W. GRIMMITT.

Ordinary Members of Council

J. BENNETT.
A. R. COOPER.
A. T. CRAWFORD, B.Sc.
C. DANNATT, O.B.E., D.Sc.
B. DONKIN, B.A.
O. W. HUMPHREYS, B.Sc.
C. R. KING, C.B.E.
H. R. L. LAMONT, Ph.D., M.A., B.Sc.
F. J. LANE, O.B.E., M.Sc.
G. S. C. LUCAS, O.B.E.
G. LYON, M.Sc.(Eng.).
SIR HAMISH D. MACLAREN, K.B.E., C.B., D.F.C., LL.D., B.Sc.
A. H. MUMFORD, O.B.E., B.Sc.(Eng.).
W. F. PARKER.
PROFESSOR M. G. SAY, Ph.D., M.Sc.
R. L. SMITH-ROSE, C.B.E., D.Sc., Ph.D.
G. O. WATSON.
J. H. WESTCOTT, B.Sc.(Eng.), Ph.D.
E. L. E. WHEATCROFT, M.A.
R. T. B. WYNN, C.B.E., M.A.

Chairmen and Past-Chairmen of Sections

Measurements:
M. WHITEHEAD.
*J. F. COALES, O.B.E., M.A.

Radio:
C. W. OATLEY, M.A., M.Sc.
*J. A. SMALE, C.B.E., A.F.C., B.Sc.

Supply:
J. D. PEATTIE, B.Sc.
*L. G. BRAZIER, Ph.D., B.Sc.

Utilization:
J. I. BERNARD, B.Sc.Tech.
*B. L. METCALF, B.Sc.(Eng.).

Chairmen and Past-Chairmen of Local Centres

East Midland Centre:
J. M. MITCHELL, B.Sc., Ph.D.
*C. D. WILKINSON.

Mersey and North Wales Centre:
P. R. DUNN, B.Sc.
*T. COATES, M.Eng.

North Midland Centre:
W. A. CROCKER.
*G. CATON.

North-Eastern Centre:
G. W. B. MITCHELL, B.A.
*H. ESTHER, B.Eng.

North-Western Centre:
PROFESSOR E. BRADSHAW, M.B.E., M.Sc.Tech., Ph.D.
*H. WEST.

Northern Ireland Centre:
MAJOR P. L. BARKER, B.Sc.
*J. R. W. MURLAND, B.Sc.(Eng.).

Western Centre
A. N. IRENS.
*J. VAUGHAN HARRIES.

Scottish Centre:
J. S. HASTIE, B.Sc.(Eng.).
*C. H. A. COLLYNS.

South Midland Centre:
A. R. BLANDFORD.
*H. J. GIBSON, B.Sc.

Southern Centre:
E. A. LOGAN, M.Sc.
*COMDR.(L) C. V. ROBINSON, R.N., O.B.E.

* Past-Chairman.

Secretary

W. K. BRASHER, C.B.E., M.A., M.I.E.E.

Deputy Secretary

F. JERVIS SMITH, M.I.E.E.

Assistant Secretary

F. C. HARRIS.

Editor-in-Chief

G. E. WILLIAMS, B.Sc.(Eng.), M.I.E.E.

THE PROCEEDINGS OF
HE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

L. 102. PART C. No. 1. MARCH 1955

DISCUSSION ON
'HE TRANSIENT RESPONSE OF R.F. AND I.F. FILTERS TO A WAVE PACKET"*

Ir. M. Dishal (*United States: communicated*): I wish to protest e strongly than Mr. D. A. Bell† against what I feel are the gerously incomplete conclusions of the author, and I feel the following comments may be helpful.

irst, let me stress that I am considering only the small-entage bandwidth problem—the problem encountered in t practical cases—with the carrier set at the middle of the band.

he author points out that eqn. (51) does not simplify when 1; this is so because the mid-frequency ω_0 of the pass band ot his ω_1 but rather is $\omega_1/\sqrt{(1-k^2)}$.

hus if eqn. (51) is correct, substitution of $u = 1/\sqrt{(1-k^2)}$ t must make its envelope reduce (for the small-percentage dwidth case) to the well-known result given by eqn. (B):

$$= 1 - \varepsilon^{-\frac{\omega_0 t}{2Q}}\sqrt{\left[1 + \frac{1}{(kQ)^2}\right]}\sin\left[\frac{\omega_0 kt}{2} + \text{arc tan}\,(kQ)\right]. \quad \text{(B)}$$

re, for the $Q_1 = Q_2$ double-tuned circuit considered by the ıor, k is the usual coefficient of coupling, and Q is the usual nant-circuit Q-factor. In the author's terminology

$$Q = 1/2s\sqrt{(1-k^2)}$$

he best way to obtain the relationship between rise-time and dwidth is to plot the complete step response from eqn. (B) not consider a part of an equation as the author has done) then relate these plots to the amplitude and phase plots for same circuit.

ig. B is such a plot of eqn. (B) for various kQ values, with response normalized to the half-power points for each case. responses have been stopped at the point where the imum overshoot occurs, and have not been plotted for values greater than 1·3, because of the resulting undesirable e overshoot, or less than 0·3, because of the resulting esirable "slowness" near the top of the step.

his Fig. B shows that one must define rise-time before it can discussed, and that if the usual 10–90% definition is used, , for *all* the amplitude and phase shapes produced by the values shown, the rise-time is approximately

$$\Delta t_{0\cdot1\to0\cdot9} = \frac{0\cdot67}{BW_{3\,\text{dB}}} \quad \text{(C)}$$

his result, which applies to the great majority of practical lications of the circuit considered in the author's Section 4, ates the sweeping conclusions drawn by him.

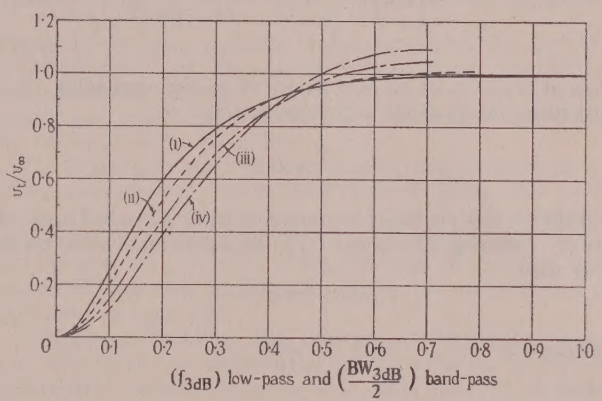


Fig. B.—Envelope response of a double-tuned circuit to a unit carrier step (carrier at centre of pass band and $\frac{BW_{3\,\text{dB}}}{f_0} < 0\cdot1$).

(i) $kQ = 0\cdot289$.
(ii) $kQ = 0\cdot576$ for maximally linear phase.
(iii) $kQ = 1\cdot0$ for maximally flat amplitude.
(iv) $kQ = 1\cdot32$.

I similarly disagree with the author's reply to Mr. Bell. If, instead of considering his P, which is based upon a part of an equation, he had plotted the actual step response, he would find that with practically useful responses there is no basis for his statement to the effect that a double-tuned circuit can be designed to give "a pulse rising just over two and a half times as fast as a pulse in a tuned-anode circuit of the same bandwidth." For the latter case it is well known that the resulting step-response envelope is for the small percentage bandwidth case

$$\frac{v_t}{v_\infty} = 1 - \varepsilon^{-\frac{\omega_0 t}{2Q}} \quad \text{(D)}$$

from which we obtain

$$\Delta t_{0\cdot1\to0\cdot9} = \frac{0\cdot70}{BW_{3\,\text{dB}}} \quad \text{(E)}$$

which for all practical purposes is identical to the result [eqn. (B)] obtained with useful double-tuned circuits.

Based upon the analysis and measurement of many higher-order physically-realizable minimum-phase circuits, it seems safe to make the following concluding statement: the 10–90% envelope rise-time of most carrier-centered small-percentage band-pass circuits is inversely proportional to its bandwidth, and

GENT, A. W.: Monograph No. 45 R, August, 1952 (see 99, Part IV, p. 326).
ee 1954, 101, Part IV, p. 164.

in fact is equal to 0·7 divided by the total 3 dB-down bandwidth, to an accuracy of approximately 20%.

Mr. A. W. Gent (*in reply*): Mr. Dishal raises the point that, as an alternative to considering the magnitude of the exponents of the decay terms for estimating the rise-time, he would prefer to take the time required for the envelope to rise from the 10% to the 90% level.

It is a simple matter to find the envelopes in either the tuned-anode case from eqn. (30), or the tuned transformer from eqn. (51). For simplicity take the tuned anode and choose the applied frequency to coincide with the resonant frequency, i.e. $u = 1$. This leads to eqn. (31), and expanding $\sin(\omega_0 t + \phi)$, this can be written

$$\frac{F(t)}{r} = p\cos\phi + q\sin\phi \quad . \quad . \quad . \quad . \quad \text{(F)}$$

where $$p = \sin\omega_1 t - \frac{\varepsilon^{-s\omega_1 t}\sin(1-s^2)^{\frac{1}{2}}\omega_1 t}{(1-s^2)^{\frac{1}{2}}}$$

$$q = \cos\omega_1 t - \varepsilon^{-s\omega_1 t}\left[\cos(1-s^2)^{\frac{1}{2}}\omega_1 t - \frac{s}{(1-s^2)^{\frac{1}{2}}}\sin(1-s^2)^{\frac{1}{2}}\omega_1 t\right]$$

. . . . (G)

Now if $y = f(x, c)$ is any family of curves depending on a parameter, the envelope is given by setting

$$\frac{\partial f}{\partial c} = 0$$

and eliminating c between this equation and the original equation for y. Applying this to eqn. (F), we differentiate with respect to ϕ, then

$$-p\sin\phi + q\cos\phi = 0$$

Solving this $$\sin\phi = \frac{q}{\sqrt{(p^2+q^2)}}$$

$$\cos\phi = \frac{p}{\sqrt{(p^2+q^2)}}$$

and substituting in eqn. (F) gives

$$\text{Envelope: } \frac{F(t)}{r} = \sqrt{(p^2+q^2)} \quad . \quad . \quad . \quad \text{(H)}$$

where p and q are given by the expressions in eqn. (G).

Exactly the same treatment can be applied to eqn. (51), but the expressions are longer and will not be given here.

If we tabulate eqn. (H) for various values of s we can find the time to rise from 10% to 90%. The 3 dB bandwidth has a simple relation to s, namely $B = 2s$. We can therefore plot the product $B\Delta t$ as a function of B. This has been done in Fig. C. The intercept at $B = 0$ is 0·70, agreeing with Mr. Dishal's eqn. (E). It will be seen that $B\Delta t$ is not constant, or even approximately constant, as B varies.

Of course, if we restrict B to small values, the product $B\Delta t$ is approximately constant, but there is nothing remarkable in that. Any function which has a Taylor expansion, and whose first derivative vanishes at some point, will be approximately constant for a restricted range of the independent variable near that point. Nevertheless, if we have a formula available which is valid for a wide range of the variable, we should surely investigate it over the whole range of its validity.

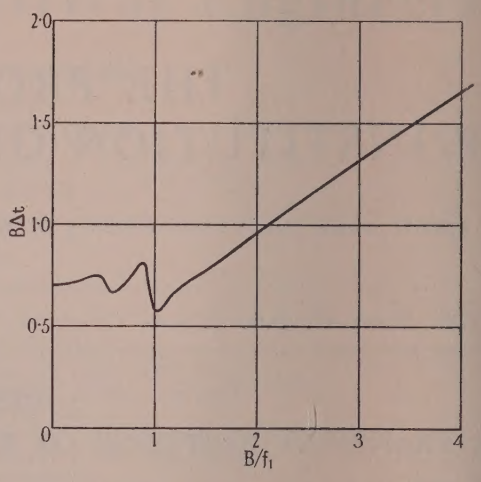


Fig. C.—$B\Delta t$ plotted against B/f_1 for a tuned-anode circuit.

In the case of the tuned transformer we have two variab so that one way of presenting the variation of $B\Delta t$ would be plot contours of constant $B\Delta t$. The actual calculation of c tours makes heavy work using the ordinary desk calculator, there is no difficulty in principle. It is a simple tabulat problem from explicit formulae. An electronic calculator wo no doubt provide a speedy result.

To show that $B\Delta t$ is not constant the calculations have b carried out for two pairs of values of m and s. In each case applied frequency was chosen to be the resonant peak frequen which differs from f_1.

(*a*) $m = 0{\cdot}9$, $s = 0{\cdot}8$. This gives peak frequency (equal applied frequency) $= 0{\cdot}745 f_1$; $B(3\,\text{dB}) = 0{\cdot}888 f_1$; $\Delta t(0{\cdot}1 \to 0$ $= 0{\cdot}748/f_1$; so that $B\Delta t = 0{\cdot}664$.

(*b*) $m = 0{\cdot}3$, $s = 10$. This gives peak frequency (equal applied frequency) $= 0{\cdot}092 f_1$; $B(3\,\text{dB}) = 0{\cdot}293 f_1$; $\Delta t(0{\cdot}1 \to 0{\cdot}9$ $4{\cdot}551/f_1$; so that $B\Delta t = 1{\cdot}333$.

These two values of $B\Delta t$ are in the ratio 2:1, so that $B\Delta$ not constant, although here again it would not be surprising find that it was nearly constant over a suitably restricted range m and s.

In conclusion, Mr. Dishal is at liberty, if he wishes, to rest his discussion to narrow bands on the grounds that these for majority of practical circuits; but although this may be t to-day, what of to-morrow? The present tendency in circ technique seems to be towards wider and wider bandwidths.

Corrigenda: 1952, **99**, Part IV, p. 334, eqn. (51).

Line 1: in the coefficient of $\sin(\omega_0 t + \phi)$, for $4ksu^2$ read $8ks^2u^2$.

Lines 2 and 3: for $\frac{1}{1-k} - \frac{u^2}{2}$ read $\frac{1}{1-k} - u^2$.

Line 4: first symbol, for + read −.

Also: for $\frac{\sin\alpha\omega_1 t}{\alpha}$ read $\frac{\sin\alpha\omega_1 t}{2\alpha s}$.

Line 5: first symbol, for − read +.

Also: for $\frac{\sin\gamma\omega_1 t}{\gamma}$ read $\frac{\sin\gamma\omega_1 t}{2\gamma s}$.

1.317.335.2

Monograph No. 100 M
May 1954

THE PRECISE MEASUREMENT OF CAPACITANCE

By J. K. WEBB, M.Sc.Eng., B.Sc.Tech., Member, and H. B. WOOD, Graduate.

(*The paper was first received* 15*th December*, 1953, *and in revised form* 20*th February*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in May*, 1954.)

SUMMARY

The method described is designed for industrial application and purpose is to enable capacitance to be conveniently measured th a precision approaching that achieved by the National Physical aboratory—on whom reliance is placed as the standardizing authority. Advantage is taken of the fact that resistance and frequency standards great stability can now be produced, and all measurements are ferred to such standards.

The apparatus essentially comprises a bridge network with induc-ely coupled ratio arms which may be switched into the form of her a "comparison" or a Wien bridge. A novel feature is the mbined use of the comparison bridge with a 2 : 1 ratio and the ien bridge with a 1 : 1 ratio, whereby the phase-angle of the fre-ency-dependent arms is made 45°. This enables the unknown pacitance in one arm to be determined in terms of the resistance in e other arm at a given frequency, the values of the other capacitance d resistance not being required. In this way capacitor standards ving values ranging from 0·01 to 1·0 μF may be precisely calibrated 1 kc/s and 2 kc/s.

The comparison bridge then enables any capacitor within the nge 100 $\mu\mu$F–1 μF to be measured in terms of one or more of these pacitor standards with an accuracy of about $\pm$100 parts in 10^6, or 0·1 $\mu\mu$F.

A simplified version of the Wien bridge is proposed which is speci-ally designed to calibrate a standard capacitor of value 0·1 μF at c/s in terms of a standard resistor having a nominal value of 80 ohms.

(1) INTRODUCTION

Mica capacitors are now being manufactured in large quan-ies, many of which are required to an accuracy of 1 000 parts 10^6 (0·1%) with correspondingly high degrees of stability d low temperature coefficients. To guarantee such limits d check their stability, a flexible system of capacitance easurement having a precision better than 100 parts in 10^6, d being capable of checking relative values within about parts in 10^6 must be available. Such precision has already en achieved in the Wien bridge arrangement of Ferguson and rtlett[3] and in Astbury's modification of the Carey–Foster idge,[6] together with the N.P.L.'s precision Schering bridge, th of which have been described by Rayner and Ford.[10]

It is very desirable, however, especially in connection with dustrial equipment, that the system of measurement should be bust and easy to operate. There is no doubt that a method ich directly compares like impedances, i.e. one capacitance with other, best satisfies this condition. Since its success depends tirely on the reliability of capacitance standards, we must st examine this question. When the present work was initiated 1945, the data available were not at all encouraging and, in 46, Garton[9] seemed to confirm this sentiment. Since then, wever, further data have accumulated and, in 1951, Rayner d Ford[14] made it appear that the position was much better an had been previously assessed. The authors have confirmed ese observations (see Section 11.6) and, furthermore, by effecting improvements in the hermetic sealing of standards, believe that they have eliminated an important source of instability. The position now is that the stability of such standards (which may be of the order of 50 parts in 10^6 per annum) can be considered adequate in those cases where annual calibrations by a standardizing authority such as the N.P.L. can readily be secured. It should be noted that only one standard, (say 0·1 μF), is necessary for the determination of any value of capacitance by the method of direct comparison, the reliability of which is therefore of a fairly high order.

One of the most successful bridges employing this method is indubitably the "capacity and conductance bridge" described by Shackleton and Ferguson.[2] A modern improvement is the substitution of inductively coupled ratio arms (see Section 11.1.5) for the original resistive ratios.

The Schering bridge is another possible choice, particularly if interest includes the measurement of power factor over a wide range. With this bridge, however, advantage cannot be taken of the facilities offered by inductively coupled ratio arms. Instead, for the most accurate work a Wagner earth is essential, which somewhat increases the complication of the network and the difficulties of operation.

The comparison bridge [Fig. 1(*b*)] may also be used to compare one capacitance with another. It has the advantage of enabling inductively coupled ratio arms to be used, although its power-factor range is restricted. If, however, interest is centred on the precise measurement of capacitance, this restriction can be accepted, since in the case of precision capacitors, power factors are usually very low. [Incidentally, the use of shunted-dial resistive elements (see Section 11.1.1), provides adequate discrimination and freedom from contact error in the bridge balance so far as power factor is concerned.]

All of the above types of bridge depend entirely for their accuracy on the stability of mica standard capacitors, and although this is shown to be good, the inherent stability of standard resistors is still considerably better. Thus the best commercial types of resistor to date have drifts lower than 10 parts in 10^6 per annum (see Section 11.7), while a recent development[16] promises drifts as low as 0·1 part in 10^6 per annum. It is therefore still attractive to be able to verify capacitance standards in terms of resistance standards.

The Wien bridge [Fig. 1(*a*)] enables this to be done, and since its components are identical with those of the comparison bridge [Fig. 1(*b*)], it can be conveniently arranged to switch from one network to the other. This resulting combination provides the means of measuring any odd values of capacitance in terms of resistance within the range 0·01–1 μF, provided that the corresponding odd values of resistance at balance can be determined with the required precision. Since this requirement is both onerous and largely superfluous, however, it is preferable to concentrate on the calibration of one or more standard capacitors in terms of fixed standard resistors, which obviates the necessity for these resistance measurements. Taken to its limit, this idea results in the very simple arrangement proposed in Section 7.

Correspondence on Monographs is invited for consideration with a view to olication.

Mr. Webb and Mr. Wood are at Standard Telecommunication Laboratories Ltd.

(2) BASIC THEORY OF THE WIEN BRIDGE

Consider the Wien bridge network shown in Fig. 1(*a*). Balance is achieved when the impedance in the A–B arm equals that in the B–C arm, or when

$$\frac{1}{\frac{1}{R}+j\omega C_1} = R_2 + \frac{1}{j\omega C_2} \quad . \quad . \quad . \quad . \quad (1)$$

from which $$C_1R_1 + C_2R_2 = C_2R_1 \quad . \quad . \quad . \quad . \quad (1a)$$

and $$\omega^2 C_1R_1C_2R_2 = 1 \quad . \quad . \quad . \quad . \quad (1b)$$

Eliminating R_2 from these equations gives

$$\frac{1}{\omega^2 R_1^2} = C_1(C_2 - C_1) \quad . \quad . \quad . \quad . \quad (2)$$

If, now, $C_1 = C_2/n$, eqn. (2) becomes

$$C_2 = \frac{n}{\sqrt{(n-1)}}\frac{1}{\omega R_1} \quad . \quad . \quad . \quad . \quad (3)$$

The optimum value of n results when the fractional change of C_2 for a small change in n is a minimum

i.e. when $$\frac{1}{C_2}\frac{dC_2}{dn} = 0$$

It can easily be shown that this condition arises when $n = 2$, in which case eqn. (3) becomes

$$C_2 = \frac{2}{\omega R_1} = \frac{1}{\pi f R_1} = \frac{G_1}{\pi f} \quad . \quad . \quad . \quad . \quad (4)$$

where $$G_1 = \frac{1}{R_1}$$

Although frequency could be made one of the variables, there are good reasons for operating at a fixed known frequency, in which case balance is secured by varying R_1 to a known value, and R_2, the value of which, however, need not be known to determine C_2.

Thus at any given frequency C_2 is given directly in terms of the effective conductance across C_1.

Substituting $2C_1$ for C_2 in eqn. (4), it follows that

$$R_1 = \frac{1}{\omega C_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Thus when $n = 2$ the phase angles of the bridge arms are 45°.

Reverting to eqn. (3), it is clear that to determine accurately an unknown capacitance C_2 precise knowledge of ω, R_1 and the function $n/\sqrt{(n-1)}$ is required.

It is, however, comparatively easy to adjust n to a value of 2 within $\pm 0{\cdot}1\%$, in which case the function has only changed by as little as $0{\cdot}5$ part in 10^6. [If $n = 2(1+\epsilon)$ where ϵ is a small fraction, then $n/\sqrt{(n-1)} \simeq 2 + \epsilon^2$.] The fact that the adjustment is not critical greatly facilitates measurements.

(2.1) Operational Procedure

To measure an unknown capacitor C_2, it is first balanced against C_1 in the comparison bridge [Fig. 1(*b*)], using a 2 : 1 ratio, so that $n = C_2/C_1 = 2$. The network is then switched into the Wien bridge [Fig. 1(*a*)] form, with 1 : 1 ratio, and balance regained simply by adjusting R_1 and R_2, the frequency being fixed at a known value. C_2 may now be determined directly in terms of R_1, and standard capacitors having values ranging

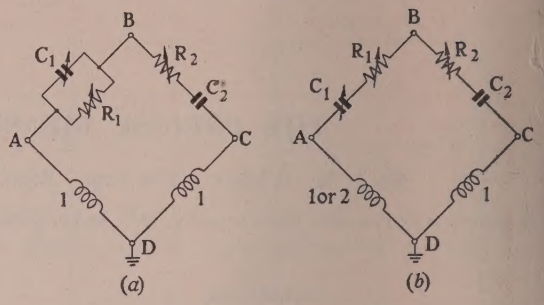


Fig. 1.—Bridge circuits

(*a*) Wien bridge of 1 : 1 ratio.
(*b*) Comparison bridge of 1 : 1 and 2 : 1 ratio. (Ratio arms are inductively coupled.)

from $0{\cdot}01$ to 1 μF may be precisely calibrated at 1 kc/s a 2 kc/s in this way. One or more of these standards then ena the variable capacitor C_1 to be precisely calibrated by the step- step procedure outlined by Ford and Astbury.[5] Capacit over the range 100 $\mu\mu$F–1 μF may thereafter be measured terms of C_1 by means of the comparison bridge using a 1:1 rat to an accuracy of ± 100 parts in 10^6 or $\pm 0{\cdot}1$ $\mu\mu$F.

The approximate values of C_1 and R_1 to enable stand capacitors of nominal values $0{\cdot}01$, $0{\cdot}1$ and 1 μF to be calibra at various frequencies, are given in Table 1.

Table 1

COMPONENT VALUES FOR WIEN BRIDGE

C_2 (unknown)	C_1	R_1 at 1 kc/s	R_1 at 2 kc/s
μF	μF	ohms	ohms
1·0	0·5	318·3	159·2
0·1	0·05	3 183	1 592
0·01	0·005	31 830	15 920

If C_1 had been made the unknown instead of C_2, the ab values of resistance would be divided by 4.

(3) CORRECTIONS

It has so far been tacitly assumed that the various cir elements are pure. Unfortunately this is not the case, an summary of the errors arising from impurities and other sou will now be considered. Those which are inherently neglig or which can be made negligible are considered first.

(3.1) Error in the Estimation of the Frequency

It is comparatively simple with present-day standards ensure that the frequency of the oscillator feeding the bri has an accuracy better than 1 part in 10^6, so that this sourc error is insignificant.

(3.2) The Ratio n

This question has already been discussed in Section 2, there is no appreciable error in the means provided for adjus n to a value of 2.

(3.3) Modulation Errors

Modulation errors have been discussed by Ferguson Bartlett.[3] They can be obviated by feeding the bridge wi very pure tone, filtering the output, and avoiding overloa elements, such as transformers and valves, which tend to bec non-linear. If the bridge balance remains undisturbed ov

ide range of input signal, the effect may be regarded as negligible. his has been achieved in the present instance.

(3.4) Variation of Effective Resistance of R_1 with Frequency

To determine C_2, the effective value of R_1 must be known at e frequency of measurement. In practice, however, R_1 is easured with direct current. At the frequencies considered kc/s and 2 kc/s) skin and proximity effects appear to be quite gligible. Conductance of the insulated covering of the anganin wire may be different for alternating current compared ith direct current, and series inductance with shunt capacitance ill contribute a real term to the a.c. impedance. From measure-ents made on similar resistors there is a definite suggestion at part, at least, of the discrepancy noted in Section 6 between sults obtained by the authors and the N.P.L. may be attributed the fact that the a.c. resistance of R_1 is anything up to about parts in 10^6 below that of the d.c. resistance, apparently used by dielectric loss. The consistency of the correlation tained over a wide range of capacitance, however, has led to e conclusion that this is the greatest error which can be tributed to such cases.

(3.5) Self-capacitance and inductance of parallel resistor R_1

Although resistors designed for use in a.c. circuits may have ry small phase-angles,[8] these are nevertheless of sufficient agnitude to produce appreciable errors in a precision bridge nless they are taken into account. Thus, while in the present se of R_1 the effect is shown to be negligible, it is later shown Section 3.12 that in the case of R_2 it is significant.

The reactance of a resistor is usually expressed in terms of an fective series inductance L_R which may be either positive or gative. An indication of the quality of a resistor is given by e time-constant, $T = L_R/R$ which should be as small as ssible. In the case of the parallel resistor R_1 [Fig. 1(*a*)] it is nvenient to express the effective reactance in terms of an uivalent parallel capacitance, C_R, since it can then be added rect to C_1 giving a total capacitance in the B–C arm of $C_1 + C_R$. his has the effect of changing the ratio n from C_2/C_1 to $_2/(C_1 + C_R)$ where $C_R = T_1/R_1$ which, in itself, is quite neg-gible as has already been explained in Section 2. The trans-rmation from the series to the parallel circuit, however, also s the effect of changing the effective value of the parallel sistor from R_1 to $R_1(1 + \omega^2 T_1^2)$. T_1 may be of the order of 5×10^{-8}, so that at 2 kc/s the factor $1 + \omega^2 T_1^2 = 1{\cdot}0000004$. s this represents a change of only 0·4 part in 10^6, the effects the self-capacitance and inductance of the parallel resistor R_1 e thus seen to be quite negligible.

(3.6) Power Factor in the Unknown Capacitor C_2

The power factor of C_2 may be expressed as $\omega C_2 r_2$, where r_2 the effective series resistance of C_2. The total effective resis-nce R_2'' in the B–C arm thus comprises the actual resistor R_2 serted in series with C_2 plus this value of r_2. Now rewriting n. (1*a*) in the form

$$\frac{C_1}{C_2} + \frac{R_2''}{R_1} = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

becomes evident that, if the ratio C_1/C_2 is adjusted to 1/2, e ratio R_2''/R_1 must also be adjusted to 1/2 to satisfy this uation. [The fact that eqn. 1(*b*) has also to be satisfied, plies that R_1 and R_2'' can each have only one discrete value.] nce, however, we have eliminated R_2'' from the final balance uations, its actual value need not be known. The power ctor of C_2 does not therefore introduce any error in computing e value of its capacitance.

(3.7) Stray Capacitances to Earth of Components in A–B and B–C Arms falling across A–D and C–D Coil Ratios

From Section 11.1.5 it is seen that large unbalanced capacitances may appear across the ratio arms without causing appreciable error. Nevertheless, the earth capacitances of the components in the A–B and B–C arms were measured and made as nearly equal as possible by connecting a small padding capacitor between the C point and earth, as indicated in Fig. 2.

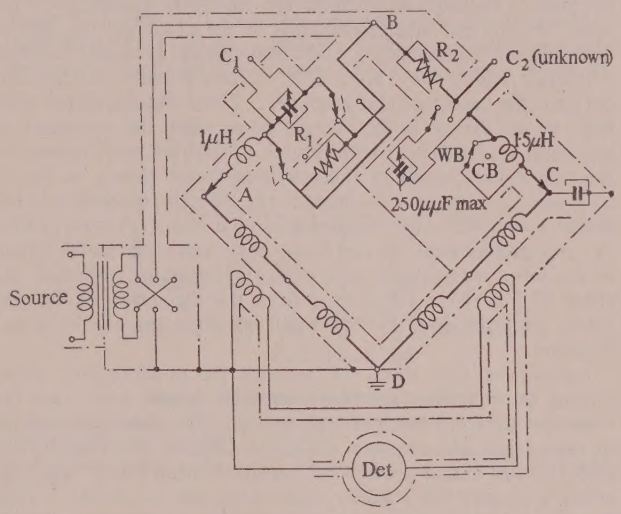


Fig. 2.—Wien and comparison bridge giving details of switching and shielding.

(Switches are shown set for Wien-bridge operation.)

A check was also made that a gross value of unbalanced capacitance (of the order of 3 000 $\mu\mu$F) could be thrown across either A–D or the C–D arms without disturbing the bridge balance. The effect of stray capacitance may thus be regarded as negligible.

(3.8) Inequality of Ratio Arms

One advantage of using inductively coupled ratio arms is that they may be adjusted to equality of ratio with great precision and, furthermore, the adjustment is very stable (see Section 11.1.5). Any slight inequality may be dealt with by the usual method of interchanging ratio arms, and to facilitate this a reversing switch has been incorporated, as indicated in Fig. 3.

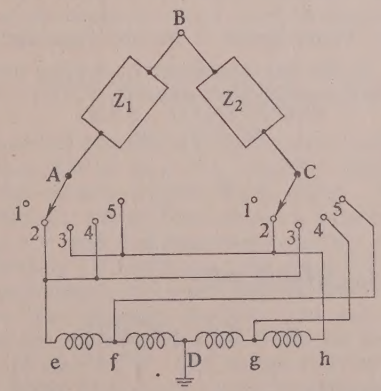


Fig. 3.—Coil-switching arrangement for Wien or comparison bridge.

1 .. Open circuit.
2 .. Equal ratio.
3 .. Equal ratio (reversed).
4 .. 1 : 2 ratio.
5 .. 1 : 2 ratio (reversed).

In the present instance, the coils have been balanced with a residual error of 7 parts in 10^6. Allowance can be made by multiplying the value of the unknown capacitance computed on the basis of equal ratio by the factor $(1 + \delta)$, where $\delta = 7 \times 10^{-6}$.

(3.9) Spurious Effects due to Direct Coupling between the Source and Detector

It is of the greatest importance to avoid any such direct coupling. Precautions include the magnetic shielding of the input transformer and tuning coils of the amplifier, the remote disposition of the oscillator, and the avoidance of mutual resistive couplings which may arise from common earth connections between the oscillator and amplifier. There should, in the ideal case, be no change in the balance point when the secondary of the input transformer to the bridge is reversed with respect to the primary by means of the reversing switch shown in Fig. 2. In practice, however, this ideal is very difficult to achieve with a bridge having very high discrimination (i.e. about ± 3 parts in 10^6), as in the present case. It was found that, with a particular setting of the input transformer and coil-ratio reversing switches, the error of 7 parts in 10^6 arising from the inequality of the coil ratios was cancelled by the small residual amount of direct coupling.

Apart from this, there still remains the possibility of errors arising from mutual couplings between bridge arms and the leads thereto from the reversing switch. The final arrangement of these leads and screen is, to quote Astbury and Ford,[6] "too often in the end a matter of hopeful compromise," and the authors cannot but ruefully agree with this sentiment. They have, however, tried various arrangements, with almost negligible effect on measured values.

(3.10) Lead Resistances and Inductances in A–B and B–C Arms

The physical connections of the components in the A–B and B–C arms will introduce some resistance and inductance in these arms. Since, however, the parallel R_1C_1 combination is electrically equivalent to the series R_2C_2 combination at balance, it is only necessary to ensure that the residual values of resistance and inductance due to leads in the A–B arm are identical with those in the B–C arm, to avoid errors on their account. A small resistive unbalance will not in itself introduce appreciable error, but it is of the greatest importance to ensure that the inductances are balanced. To this end one arm is "padded" with sufficient inductance (of the order of 1 μH) to attain this balance by the method outlined in Section 11.2.

(3.11) Power Factor of Parallel Capacitor C_1

From eqn. (4) the test capacitance C_2 is given direct in terms of the effective conductance G_1' across C_1.

Now $$G_1' = G_1 + g_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where $$G_1 = \frac{1}{R_1}$$

and $$g_1 = \omega C_1 \phi_1$$

ϕ_1 being the power factor of C_1.

Hence $$G_1' = G_1\left(1 + \frac{\omega C_1 \phi_1}{G_1}\right)$$

But, from eqn. (5)

$$\omega C_1 R_1 = 1$$

Thus $$\frac{1}{R_1'} = \frac{1}{R_1}(1 + \phi_1) \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

In practice C_1 comprises a high-grade mica capacitor hav a power factor ranging from about 0·00004 to 0·0002, depend on the setting. If, by the way of example, the power facto 0·0001, it is seen that $1/R_1$ is increased by a factor of 100 pa in 10^6, which is a significant correction.

A careful check must therefore be kept on the value of power factor of C_1, as mentioned in Section 11.1.3.

(3.12) Self-Capacitance and Inductance of Series Resistor R

Although the actual resistance of R_2 need not be kno its phase angle is important, since it affects the total reacta in the B–C arm and hence the apparent value of the sam capacitance.

Thus, if R_2 has an effective series inductance L_2, such t $T_2 = L_2/R_2 \simeq 2L_2/R_1$, the reactance in the B–C arm becom

$$X_{BC} = \frac{j\omega R_1 T_2}{2} + \frac{1}{j\omega C_2}$$

Now from eqn. (4) $$\omega R_1 = \frac{2}{C_2}$$

so that $$X_{BC} = \frac{jT_2}{C_2} + \frac{1}{j\omega C_2}$$

$$= \frac{1 - \omega T_2}{j\omega C_2} = \frac{1 - Q_2}{j\omega C_2}$$

where $$Q_2 = \omega T_2$$

But $$X_{BC} = \frac{1}{j\omega C_{BC}}$$

where C_{BC} is the effective capacitance in B–C arm.

Hence $$C_2 = C_{BC}(1 - Q_2) \quad . \quad . \quad . \quad .$$

Or, to find the true value of C_2, we must multiply the appar measured value by the factor $(1 - Q_2)$.

Table 2 gives the order of the correction. In those ca where Q_2 is negative, $(1 - Q_2)$ is greater than unity.

Table 2

CORRECTION DUE TO TIME-CONSTANT OF R_2

C_2	Correction due to time-constant of R_2	
	Parts in 10^6 at 1 kc/s	Parts in 10^6 at 2 kc/
μF		
1	0	0
0·1	−10	−11
0·01	−209	−630

(3.13) Residual Capacitance C_0 in the B–C Arm

It was shown in Section 2 that the Wien-bridge equati give the total reactance in the B–C arm. In order, therefore extract the value of the capacitance connected to the test minals of the bridge from the measured data, it is necessar know the internal residual capacitance of the B–C arm. was determined using the bridge network itself as follows:

The comparison bridge is first balanced with a conven fixed value of capacitance C_2 connected to the "unkno terminals in the B–C arm with the coils switched to unity ra The capacitance in the A–B arm is then $C_1' = C_2 + C_0$, w C_0 is the required residual capacitance. When the coils are switched to a ratio of 2 : 1, the capacitance in the A–B arr balance is adjusted to $C_1'' = \frac{1}{2}(C_2 + C_0)$. From these

quations $C_0 = 2(C_1' - C_1'') - C_2$. Results were found to be eproducible over an extended period to better than 0·1 $\mu\mu$F, he mean value being 25·2 $\mu\mu$F. This residual must be subracted from all Wien-bridge readings in order to determine the alue of the capacitance actually connected to the test terminals.

3.14) External Lead Capacitance C_0' and Inductance connecting Unknown Capacitance to Test Terminals

In addition to the fixed value of the residual capacitance in he B–C arm, an allowance must be made for the capacitance C_0' f the connecting leads. This external capacitance, which should lways be kept as small as possible, can likewise be determined y means of the comparison bridge. The relative effect of direct nd mutual capacitance must be carefully watched, however, s pointed out by Field.[4] If short connecting leads are used, a sual value for the external lead capacitance C_0' is about 0·8 $\mu\mu$F, o that the total allowance for residual and external leads amounts o 26·0 $\mu\mu$F. The uncertainty in this value may be of the rder of 0·1 $\mu\mu$F and, although this is quite negligible when easuring capacitances of 0·1 μF and over, it may involve an rror of as much as 100 parts in 10^6 when measuring a 1 000-$\mu\mu$F apacitor.

As indicated in Section 11.2, the inductance of the connecting eads should be kept as low as possible, and an appropriate orrection made.

(4) COMPUTATION OF CORRECTIONS

The success of any practical scheme of measurement largely epends on the ability to take due account of unavoidable orrection terms without the necessity of too much computation, nd in what follows it is demonstrated how this can be achieved n the present instance.

Taking the significant correction factors into account, eqn. (4) ust be rewritten as

$$C_2 = \frac{1}{\pi f R_1}(1 + \phi_1)(1 - Q_2) - (C_0 + C_0') \quad . \quad (10)$$

nd, since ϕ_1 and Q_2 are both much smaller than unity, this losely approximates to

$$C_2 = \frac{1}{\pi f R_1}(1 + \phi_1 - Q_2) - (C_0 + C_0') \quad . \quad . \quad (11)$$

hich is now in a form suitable for computation.

(4.1) Illustrative Example

By way of example, a set of observations appertaining to the alibration of a 0·1-μF standard capacitor at 1 kc/s with the resent Wien bridge is now given, which will serve to illustrate he procedure.

The measurement is made in terms of a nominal 3 183-ohm esistance standard, the exact value of which is known from N.P.L. calibrations. The comparison bridge with a 2 : 1 ratio s first used to set C_1 so that $C_2/C_1 = 2$. A 6-dial Wheatstone ridge is used to compare the actual value of the variable resistor R_1 in the Wien bridge at balance with this 3 183-ohm standard.

In this way we obtain:

$R_1 = 3\,185{\cdot}26$ ohms
$\phi_1 = 0{\cdot}000154$
$Q_2 = -0{\cdot}00001$
$C_0 = 0{\cdot}0000252$ (see Section 3.13)
$C_0' = 0{\cdot}0000008$ (see Section 3.14)

From which $C_2 = 0{\cdot}0999225$ μF at 19·1° C.
Corresponding N.P.L. value $= 0{\cdot}0999260$ μF at 19·1° C.

The difference is 35 parts in 10^6 (for a discussion on the significance of this, see Section 6).

(5) CONSISTENCY OF RESULTS

A decade mica-capacitor was calibrated on the comparison bridge throughout the range 0·01–1·0 μF by the step-by-step method. The calibration was standardized in terms of a Wien-bridge measurement of the 0·1-μF setting. Each setting of the capacitor was then remeasured, using the Wien bridge in conjunction with a precision 6-dial Wheatstone bridge calibrated in terms of 100-, 1 000- and 10 000-ohm standards, and the results were compared. Measurements were made at both 1 kc/s and 2 kc/s. This whole procedure has been repeated at intervals on four separate occasions. The range of the deviations between all such two sets of observations is indicated in Fig. 4.

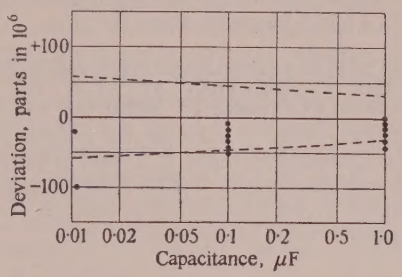


Fig. 4.—Correlation of observations.

The dotted lines give the range of the deviations between Wien- and comparison-bridge measurements as detailed in Section 5.
The points record the differences between the authors' measurements on their Wien bridge and the N.P.L. measurements of several capacitor standards, as described in Section 6.

All measurements were made in a laboratory which is normally subject to a temperature variation of several degrees in the course of a day. As the time interval between any two sets of measurements on any particular setting of the capacitor was about four hours, part of the range must be attributed to actual changes due to its temperature coefficient. It is estimated that if such changes were eliminated the range would be reduced almost by 50%.

(5.1) Measurement of the Stability of a Capacitor

In measuring the stability of a capacitor we are not concerned with absolute values, but only with values relative to some fixed and invariable standard. From the foregoing considerations, it is seen that the range arising from experimental error should hardly amount to more than about ±20 parts in 10^6, but it must be emphasized that extreme care, particularly in the matter of temperature control, must be exercised to achieve this accuracy.

(6) CORRELATION OF WIEN-BRIDGE RESULTS WITH N.P.L. CALIBRATIONS

Several standard capacitors, including those listed in Table 3, were measured first by the N.P.L. and then immediately afterwards by the authors on their Wien bridge. The difference between these two sets of observations is indicated on Fig. 4, from which it is seen that there is an apparent discrepancy of about 30 parts in 10^6 throughout the range 0·01–1·0 μF. It must be remembered, however, that, at best, capacitance is only known to an accuracy of about ±50 parts in 10^6, while results normally given by the N.P.L. have the estimated limits of error of 100 parts in 10^6. For routine measurements it is not possible to justify limits much closer than this, but in the case of the authors' standards, by special arrangement the N.P.L. authorities agreed to record the values they obtained to the nearest 10 parts in 10^6, so that "rounding off" should not introduce any error and as much information as possible, parti-

cularly as regards stability, would be given. Against this background, the discrepancy can be regarded as being almost negligible, and it becomes apparent only from statistical evidence. The fact that the ratio of a.c. to d.c. resistance of R_1 may be slightly less than unity may account for part, at least, of the discrepancy, as mentioned in Section 3.4, but this requires further investigation. For commercial purposes it is sufficient to be able to predict N.P.L. results with certainty from bridge readings, and in the present instance this can be achieved by multiplying all values by a constant factor of 1·00003. It may be concluded, however, that the various correction terms have been accurately determined and applied, since none of the terms involved would give rise to a constant percentage error throughout a capacitance range of 100 : 1 and a frequency range of 2 : 1.

(7) PROPOSALS FOR A SIMPLIFIED WIEN BRIDGE DESIGNED FOR THE CALIBRATION AT 1 KC/S OF A FIXED CAPACITOR STANDARD OF NOMINAL VALUE 0·1 μF

If a capacitance bridge is available, the internal standards of which can be calibrated in terms of a fixed reference standard capacitor, a simplified form of Wien bridge designed specifically for the calibration of this standard can be adopted. It is generally agreed that the best value to choose is 0·1 μF.

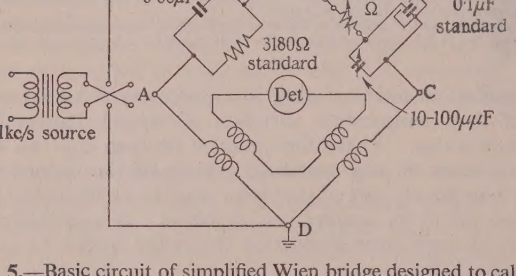


Fig. 5.—Basic circuit of simplified Wien bridge designed to calibrate a 0·1-μF standard capacitor at 1 kc/s.

Referring to Fig. 5, the scheme is to balance a 0·1-μF standard connected in parallel with a small built-in variable air-capacitor of range about 10–100 μμF against a standard resistor of nominal resistance 3 180 ohms, whose exact value is initially adjusted so that balance is secured at about mid-scale of the variable air-capacitor. Variations of this air capacitor will not affect the value of the ratio C_1/C_2 by more than 0·1%, so that, as shown in Section 2, the resulting error in the computation of C_2 using eqn. (11) is less than 1 part in 10^6. In series with the 0·1-μF standard is a fixed resistor of about 1 585 ohms, together with a decade resistor of the constant-inductance type, having a total resistance of 10 ohms adjustable in increments of 0·001 ohm, the last two decades being of the shunted-dial type (see Section 11.1.1). The nominal 0·05-μF capacitor in the A–B arm is adjusted to be equal to half the total capacitance in the B–C arm to within about ± 0·1% when the setting of the air capacitor is about mid-scale. The coils are designed to give ratios of 1 : 1 or 2 : 1 by switching. The screening of the bridge is similar to that indicated in Fig. 2, and such adjustments and corrections must be made as have already been described.

Apart from the calibration of the 3 180-ohm resistor, this scheme entirely obviates the need for measurements of resistance, which is an advantage. The value of the 0·1-μF standard is determined in terms of the value of the effective conductance in the A–B arm and, provided that the nominal 3 180-ohm resistor and the power factor of the 0·05-μF capacitor are quite stable, the stability of the 0·1-μF standard is indicated directly from observations over a period of time of the settings of the variable air-capacitor required to achieve balance. Notes concerning the construction of this 3 180-ohm resistor are given in Section 11. The question of the power factor of the parallel 0·05-μF capacitor, which is a second-order correction, has been discussed in Section 3.11.

It can be assumed that both the temperature and frequency coefficients of the 0·1-μF standard remain stable with time, so that these can be initially measured independently.

(8) CONCLUSIONS

The Wien–comparison-bridge network described provides a convenient and flexible method well suited to industrial application of measuring capacitance within the range 100 μμF to 1 μF at audio frequencies. A precision of ±100 parts in 10^6 or ±0·1 μμF can be achieved, whilst relative values can be determined within limits of ±20 parts in 10^6.

(9) ACKNOWLEDGMENTS

The authors gratefully acknowledge help which they have received from their colleagues, Messrs. R. M. Barnard and C. H. Windsor, and the co-operation of the N.P.L. authorities in particular Messrs. G. H. Rayner, L. H. Ford and E. J. Tomlin. They would also like to acknowledge assistance by Messrs. H. W. Sullivan Ltd., and H. Tinsley and Co., in supplying components, and to Mr. D. C. Gall for some helpful discussion on the design of resistance standards.

The paper is published by kind permission of Standard Telecommunication Laboratories Ltd.

(10) BIBLIOGRAPHY

(1) MUELLER, E. F.: "Wheatstone Bridges for Resistance Thermometry," Bulletin of the National Bureau of Standards, 1916–17, p. 547.

(2) SHACKELTON, W. J., and FERGUSON, J. C.: "Electric Measurement of Communication Apparatus," *Bell System Technical Journal*, 1928, **7**, p. 70.

(3) FERGUSON, J. C., and BARTLETT, B. W.: "The Measurement of Capacitance in Terms of Resistance and Frequency," *ibid.*, p. 420.

(4) FIELD, R. F.: "Connection Error in Capacitance Measurements," *General Radio Experimenter*, 1938, **12**, No. 8, p.

(5) FORD, L. H., and ASTBURY, N. F.: "A Note on the Calibration of Decade Condensers," *Journal of Scientific Instruments*, 1938, **15**, p. 122.

(6) FORD, L. H., and ASTBURY, N. F.: "The Precision Measurement of Capacitance," *Proceedings of the Physical Society*, 1939, **51**, p. 37.

(7) GOLDING, E. W.: "Electrical Measurements and Measuring Instruments" (Pitman, London, 1940).

(8) HAGUE, B.: "Alternating Current Bridge Methods" (Pitman, London, 1945).

(9) GARTON, C. G.: "The Characteristics and Error of Capacitors used for Measurement Purposes," *Journal I.E.E.*, 1946, **93**, Part II, p. 397.

(10) RAYNER, G. H., and FORD, L. H.; "The Calibration of Capacitors at the National Physical Laboratory," *ibid.*, 1948, **95**, Part II, p. 312.

(11) FORD, L. H.: "The Effect of Humidity on the Calibration of Precision Air Capacitors," *ibid.*, p. 13.

(12) CLARK, H. A. M., and VANDERLYN, P. B.: "Double-range A.C. Bridges with Inductively Coupled Ratio Arms," *Proceedings I.E.E.*, Paper No. 742 M, November, 1949, (**96**, Part II, p. 365).

(13) WOOD, H. B.: "A New Amplitude-Stabilized Variable Frequency Oscillator." Standard Telecommunication Laboratories Ltd., Technical Report No. 38, August, 19

14) RAYNER, G. H., and FORD, L. H.: "The Stability of Mica Standards of Capacitance," *Journal of Scientific Instruments*, 1951, **28**, p. 168.
15) FELTON, A.: "The Accurate Measurement of Electrical Standards," *Proceedings I.E.E.*, Paper No. 1153 M, May, 1951 (**98**, Part II, p. 694).
16) BARBER, C. R., GRIDLEY, A., and HALL, J. A.: "A Design for Standard Resistance Coils," *Journal of Scientific Instruments*, 1952, **29**, p. 65.
17) WOOD, H. B., and WINDSOR, C. H.: "A Very Sensitive Audio-Frequency Bridge Detector Amplifier." Standard Telecommunication Laboratories, Ltd., Technical Report No. 50, July, 1952.
18) HARRIS, F. K.: "Electrical Measurements" (Chapman and Hall, London, 1952).
19) HARTSHORN, L.: "Units and Standards of Electrical Measurement," *Proceedings I.E.E.*, Paper No. 1403, November, 1952 (**99**, Part I, p. 271).
20) RAYNER, G. H.: "A Selective Detector Amplifier for 10–10 000 c/s," *Journal of Scientific Instruments*, 1953, **30**, p. 17.
21) RAYNER, G. H., and FORD, L. H.: "The Performance of Dried and Sealed Mica Capacitors," *ibid.*, 1954, **31**, p. 3.

(11) APPENDICES

(11.1) Notes concerning Apparatus

1.1.1) Resistor R_1.

The resistor R_1 has eight decades and is variable in increments f 0·001 ohm from about 0·37 ohm to about 39 999·99(10) ohms. nly three 10 000-ohm coils are present, which minimizes the alue of the capacitance of the coils to the screen. The time-onstants of R_1 are of the same order as those of R_2. The last ecade is of the shunted-dial type, which provides the means of btaining very small increments of resistance without introducing ppreciable errors arising from switch contacts, provided a esidual value of resistance can be tolerated. A series of decades f such resistors may be readily computed by the ingenious ethod given by Mueller.[1] There was some doubt whether ccuracy would be maintained at 1 or 2 k/cs owing to the effect f the relative time-constants of the parallel paths, but this doubt as dispelled following measurements made at 1 kc/s on a ecially constructed decade which gave increments of 0·001ohm. his device had been known for nearly 50 years, and considering ie facilities which it offers, it is surprising that it has not become ore popular.

1.1.2) Resistor R_2.

The resistor R_2 is similar in construction to R_1 but it has only even decades, variable in increments of 0·01 ohm from zero to 9 999·99 ohms. Since the time-constant of this resistor may troduce a significant correction term (see Section 3.12) the alues were determined by the N.P.L. at 1 kc/s and 20° C, om which the values of reactance at 1 kc/s and 2 kc/s were educed.

1.1.3) Capacitor C_1.

The capacitor C_1 has five dials and is continuously variable om its residual value of 182·2 $\mu\mu$F to about 1·0111 μF. The st dial is a variable air capacitor, while the next one is also an r capacitor having nine 100-$\mu\mu$F steps. The remaining decades se best-quality mica. Two terminals protrude from the bridge enable capacitors to be connected in parallel with C_1.

Since the power factor of C_1 represents a significant correction ee Section 3.11), it becomes necessary for it to be determined any given setting. This is best done from measurements of conductance by the step-by-step method.[5] The tabulated values of conductance enable the power factor at any setting to be computed readily.

(11.1.4) Zero-balance capacitor.

The function of the zero-balance capacitor is to balance out the residual capacitance of C_1 when using the comparison bridge. It comprises a variable air-capacitor of maximum capacitance 250 $\mu\mu$F in parallel with a 150-$\mu\mu$F mica capacitor.

(11.1.5) Inductive ratio arms.

For those bridge networks to which it is applicable, the tightly-coupled inductive ratio-arm offers many advantages, including the relative ease of initial adjustment to a high degree of accuracy, great stability, and low effective impedance at balance. Essential design features have been summarized by Hague,[8] Clark and Vanderlyne,[12] and Harris[18] and it is sufficient to note that, when the bridge is balanced, the two windings carry currents such that the m.m.f.'s associated with the windings are of equal magnitude and in opposite sense. Under these conditions the windings become virtually non-reactive, so that the effective impedance comprises only the resistance and leakage reactance which, in the present instance, amounts to $5 + j9{\cdot}5$ ohms at 1 kc/s, and $5 + j19$ ohms at 2 kc/s. (The self-inductance is almost 5H and the coupling coefficient is about 99·97%.)

To achieve a 1 : 1 or, alternatively, a 2 : 1 ratio, a Permalloy C core is wound with star-quadded wire and connections made as indicated in Figs. 3 and 6. Although this form of

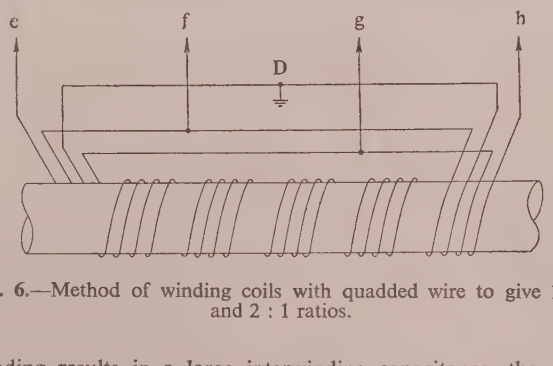


Fig. 6.—Method of winding coils with quadded wire to give 1 : 1 and 2 : 1 ratios.

winding results in a large interwinding capacitance, the only effect is a slight reduction in bridge sensitivity, which in any case is quite adequate.

The detector coil is wound symmetrically over the ratio-arm windings, from which it is electrically screened by means of a copper shield. The complete coil is magnetically screened by enclosing it in a Permalloy box.

(11.1.6) Switches.

The switching arrangements of the bridge networks are shown in Figs. 2 and 3. There are no special requirements for most of these switches, apart from the usual considerations of screening in order to avoid coupling between contacts, rotors, etc., and ordinary commercially available types of switch have consequently been employed.

The only case in which a high-quality switch is essential is that of the series–parallel device in the A–B arm of the bridge. This switch must have a very low and constant resistance, because it is included in the measurement of R_1. The design adopted comprises a number of rhodium-plated inserts. This switch forms an integral part of the internal capacitor C_1.

(11.2) Method of balancing out Residual Lead Resistances and Inductances in the A–B and B–C Arms of the Wien and Comparison Bridges

With reference to Fig. 7, the object of the Wien-bridge balance is to equate the impedance of the parallel circuit comprising

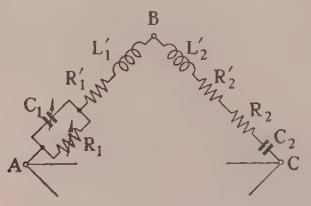


Fig. 7.—Equal-ratio Wien bridge with residual lead resistance R_1R_2' and inductance $L_1'L_2'$.

C_1R_1 with that of the series circuit C_2R_2. To avoid the necessity for correcting for residual lead resistances and inductances, it is consequently desirable to make $R_1' = R_2'$ and $L_1' = L_2'$.

The resistive balance (which is not critical) can readily be adjusted by means of d.c. measurements to within limits which involve negligible error. The inductive balance, however, requires very precise adjustment, an unbalance of less than $0{\cdot}1\ \mu$H being desirable. The method of effecting this balance is explained by reference to Fig. 8.

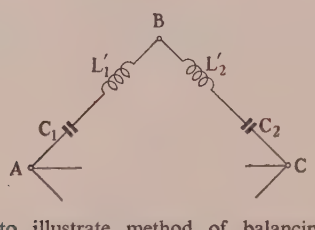


Fig. 8.—Circuit to illustrate method of balancing out residuals L_1' and L_2'.

Neglect all resistances and stray capacitances: then $C_1 = C_2$ if $L_1' = L_2'$.

Thus, by comparing C_1 and C_2 on an auxiliary bridge, L_1' can be adjusted to equal L_2'. Such a direct comparison, however, would be invalidated by the effect of stray capacitances, but this difficulty can be overcome by the following device.

The bridge is first balanced with C_1 and C_2 made as low as convenient. A large fixed capacitance ΔC_2 is then added to C_2 and the bridge rebalanced by increasing C_1 by ΔC_1, which is compared with ΔC_2 by means of the auxiliary bridge. Inductance is then added to one or other bridge arm until $\Delta C_1 = \Delta C_2$.

As a practical example, if $\Delta C_2 = 0{\cdot}8\ \mu$F, then, for an unbalance of $\pm 0{\cdot}1\ \mu$H, $\Delta C_1 = 0{\cdot}8\ \mu\text{F} \pm 10\ \mu\mu$F at 2 kc/s. An auxiliary capacitance and conductance bridge, of the type described in Reference 2, is available, on which capacitances of this order can be compared to within about 1 $\mu\mu$F.

To accommodate any difference of power factor between C_1 and C_2, R_2 may be inserted in the B–C arm. This resistor, which has eight decades, can be varied in increments of 0·001 ohm. Its last five decades are of constant inductance ($1{\cdot}5\ \mu$F approximately), which may be regarded as part of L_2'.

If R_2 should tend to negative values, a resistor (probably less than 1 ohm) can temporarily be inserted in series with C_1 in the A–B arm.

When R_1 is switched in parallel with C_1, the complete circuit of the A–B arm is given in Fig. 9.

C_1 is connected to R_1 by means of leads of inductance L_3'

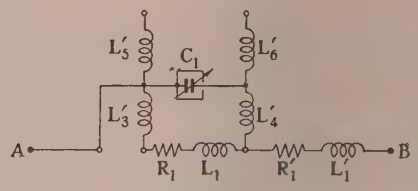


Fig. 9.—Equivalent A–B arms of Wien bridge with switch in "parallel" position.

and L_4' and to the available top-panel terminals through lea of inductances L_5' and L_6'. The value of any one of these indu tances is less than $0{\cdot}25\ \mu$H.

In the Wien-bridge balance L_5' and L_6' are not in circu while L_3', L_4' and L_1, the inductance of R_1, have been shown Section 3.5 to have negligible effect. Hence the only lea inductance to be considered is L_1'. If, however, R_1 is disconnect from the sub-panel terminals (without changing the serie parallel switch position), L_4' is now effectively in series with so that its apparent reactance is $\left(\omega L_4' - \frac{1}{\omega C_1}\right)$. In measuring by means of the auxiliary bridge, however, if connections a made to the top-panel terminals, its apparent reactance $\left[\omega(L_5' + L_6') - \frac{1}{\omega C_1}\right]$. If, then, we contrive to make L_4' $(L_5' + L_6')$, the effect of L_4 is eliminated.

Although L_4', L_5' and L_6' cannot be measured individuall their values can be deduced to an estimated limit of $0{\cdot}05\ \mu$ from measurements of various combinations with L_3', and fro considerations of the geometry of the layout. The resul obtained are $L_4' = L_5' + L_6' = 0{\cdot}23\ \mu$H.

When the series–parallel switch is set to the series positio the bridge is transformed into a capacitance ratio bridge and t circuit of the A–B arm is as shown in Fig. 10. The lead indu

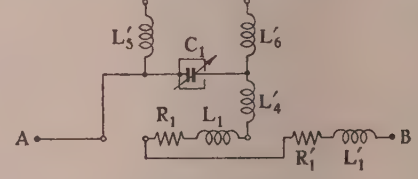


Fig. 10.—Equivalent A–B arm of comparison bridge with switch in "series" position.

In practice it is arranged that $L_4' \simeq L_5' + L_6' \simeq 0{\cdot}23\ \mu$H.

tance L_1' is now increased by $L_1 + L_4'$ and a like increment must inserted in the B–C arm to preserve balance. This adjustm can be effected by employing the same technique as alre described. The effect of $L_5' + L_6'$ on the apparent value of as measured by the auxiliary bridge must now be coun balanced by like inductances connected to C_2.

This extra increment of inductance only appears in the A arm when the series-parallel switch is in the series or comparis bridge position. In the parallel or Wien-bridge position i short-circuited by a supplementary switch, as indicated in Fig

(11.3) Standard-Frequency Source

Many laboratories have adopted the scheme of generatin standard frequency—usually 1 kc/s—and distributing the si through buffer stages to various locations. The source is usu controlled with very great precision (often to a few parts in 1 There is the danger, however, of introducing some phase m lation arising out of disturbances in transmission. The met

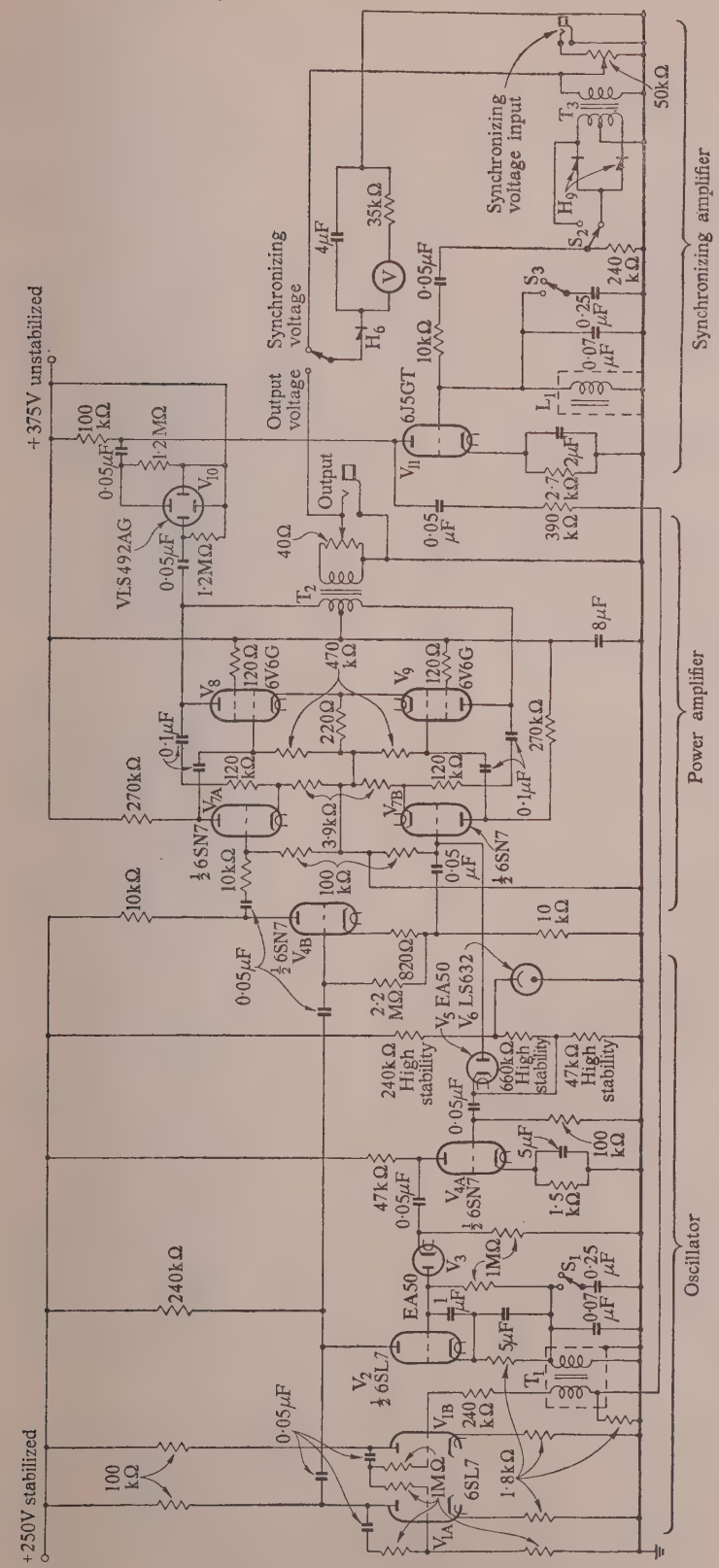


Fig. 11.—Bridge oscillator circuit.

S1 S2 and S3: Ganged switches.
Transformer T_1: Ratio 2:1.
Secondary inductance 100 mH.
Q-factor 50.
Transformer T_2: Ratio 18:1.
Primary inductance 13 H.
Inductance L_1 = 100 mH.
Q-factor 100.
H_6 and H_9: Miniature selenium rectifiers.

adopted by the authors, of locking a free-running stable oscillator as lightly as possible to the 1-kc/s signal, largely obviates this danger.

Since an accuracy in frequency of 1 part in 10^6 would be adequate, an alternative practical scheme would be a bridge oscillator controlled by an internal crystal.

(11.4) The Bridge Oscillator

The essential requirements of a suitable Wien-bridge oscillator are as follows:

(*a*) An output of about 10 volts with an impedance of about 40 ohms. (In practice the oscillator is operated at a level of about 1 volt.)

(*b*) The harmonic content of the output to be less than 0·3%. (Filters inserted between oscillator and bridge, which reduced this content by 70 db, were found to have no effect on either the bridge balance or its discrimination.)

(*c*) Convenient means of generating frequencies of 1 kc/s and 2 kc/s which can be locked to a 1-kc/s standard-frequency source, synchronism being observed from a Lissajous figure on a miniature built-in cathode-ray tube.

(*d*) The intrinsic frequency stability to be of a high order, so that only very light locking is necessary. This helps to obviate phase modulation and noise, inherent in the 1-kc/s standard source.

The circuit of an oscillator which meets this specification is given in Fig. 11. This circuit is based on a more general design, details of which are given in Reference 13.

(11.5) Bridge Detector Amplifier

The authors used the detector unit described in Reference 17, but an alternative design which should prove equally suitable has recently been published by Rayner.[20]

(11.6) Factors associated with Standard Capacitors

The stability of mica standards of capacitance has recently been examined by Rayner and Ford,[14] and the authors can confirm that at least some of their own standards have given similar high orders of stability. These are certainly better than was believed possible when the present investigation was started in 1945, and the only data available were those given by Garton.[9] Table 3 indicates the magnitude of these drifts.

Capacitors Nos. 1, 2 and 3 are of a well-established commercial type, and although the manufacturers say that they are "sealed in air-tight metal cases" the authors cannot agree that this statement is strictly correct. It soon became evident that such capacitors were affected by humidity, and the variations observed are believed to be mainly due to this cause. The deleterious effect of humidity has been confirmed by Rayner and Ford,[14] who have also noted the beneficial effect on the stability of very thorough drying.[21] The authors accordingly arranged with their own organization to have standards Nos. 4 and 5 constructed, which are truly hermetically sealed. It is too early yet to pass final judgment on the performance of the latter type, but the results to date are at least promising.

Standard capacitors are usually deficient in so far that they have no facility for the insertion of a thermometer. To improve

Table 3

DRIFTS OF CAPACITOR STANDARDS

Calibration date	Drifts deduced from N.P.L. calibrations Standard No. 1	2	3	4	5
	1 μF	0·1 μF	0·01 μF	1 μF	0·1 μ
	Parts in 10^6 from the initial value				
26. 6.1946	0				
11. 9.1946			0		
16. 9.1949	+10		+1 020		
31. 8.1950	+20	0	+110		
7. 2.1951	+50	+20	+310		
8.10.1951				0	0
11.10.1952	+40	+130	+1 860	+20	+2
11. 2.1953				+10	−1
Temperature coefficient, parts in 10^6 per deg C (range 20–25° C) at 1 kc/s	−8	−4	+28	+14	+2

matters the authors have encased their hermetically seal standards Nos. 4 and 5 (which are rectangular in shape) in heavy brass box having walls 0·5 in thick. The box has detachable top and is lagged with cork 0·625 in thick: a verti hole drilled into its side enables a thermometer to be insert This greatly facilitates accurate observations of temperature.

(11.7) Factors Associated with Standard Resistors

Every resistor is something of a strain gauge and consequen more or less unstable. Progress is still being made in the p duction of strain-free resistors of high stability. The situati so far as the N.P.L. and the N.B.S. are concerned has recen been recounted by Hartshorn,[19] from which it appears that art of constructing stable strain-free hermetically-sealed resist has reached a very high level of perfection, drifts on the b coils being only about 1 part in 10^7 per annum. Great skill design and construction, however, is necessary to achieve s results.

From annual N.P.L. measurements made on three commer standard resistors having values 100, 1 000 and 10 000 oh over a period of five years, the maximum deviation from initial value was found to be only 30 parts in 10^6. These resist are oil-immersed and in canisters, but they are not sealed, while the results are reassuring, improvements in constructi including hermetic sealing, should reduce the drifts still furt Alternative designs,[15,16] which promise almost zero drift, now also becoming available.

From Sections 3.4 and 7, it should be noted that in the c of a resistor such as $R_1 = 3\,183$ ohms, it is necessary to k the a.c. resistance at 1 kc/s, which suggests the desirability making the ratio of a.c. to d.c. resistance as nearly unity possible. The time-constant, so long as it is reasonably low of minor importance. If a resistor of the requisite high stab for this a.c. application cannot be produced, it will be neces to check R_1 with direct current against an external stand resistor designed specifically for high stability, without re to its time-constant.

21.3.015.5 : 621.3.011.21

Monograph No. 101
June 1954

THE A.C. IMPEDANCE OF PLASMA DISCHARGES IN MERCURY VAPOUR

By S. E. YUSSUF, Ph.D., Graduate, and Professor J. C. PRESCOTT, D.Eng., Member.

(*The paper was first received 3rd September*, 1953, *and in revised form 9th March*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in June*, 1954.)

SUMMARY

Whenever an alternating electric field of small amplitude is applied to a d.c.-sustained plasma discharge, a corresponding alternating component will appear in the discharge current. The discharge path behaves as a conductor of inductive impedance; the inductance and resistance depend on the frequency of the superimposed quantities.

The paper deals with the effect of the behaviour of electron and thermal ionization on the impedance of the low- and high-pressure positive columns of a plasma discharge in mercury vapour. The particular characteristics of the positive column in each case are outlined and an expression for the effective impedance per unit length of the column is derived. Typical theoretical and experimental curves of the variation of both components with frequency are given for the low-pressure case. For the high-pressure column the theoretical curves are of a shape similar to those obtained experimentally by Magnold for an arc discharge in mercury vapour. Although Magnold did not give sufficient particulars of the arc discharge investigated for an exact comparison to be made, the theory seems to explain the physical phenomena that take place in the high-pressure positive column.

(1) LIST OF SYMBOLS

p = Gas pressure, baryes.
T_g = Temperature of gas molecules, °K.
T = Electron temperature, °K.
e = Electron charge, coulombs.
V_i = Ionization potential of mercury vapour, electron-volts.
k = Boltzmann's constant.
α = Rate of production of ions, ions/electron/sec.
D = Coefficient of ambipolar diffusion.
N = Electron concentration per cubic centimetre in the positive column of the d.c. arc.
η = Total number of electrons per centimetre length of the positive column.
I = Direct arc current, amp.
$\mathscr{E}$ = Longitudinal electric field in the d.c. column, volts/cm.
V_p = Average thermal energy of positive ions.
λ = Electron mean free path, cm.
m = Electron mass, g.
μ = Electron mobility, cm/sec/volt/cm.
$N = n$ = Amplitude of sinusoidal change in electron concentration, per cubic centimetre.
v_0 = Mean electron velocity, cm/sec.
a = Radius of the tube container in the low-pressure column and effective radius of high-pressure column, cm.
N_g = Concentration per cubic centimetre of neutral gas molecules.
α_1 = Coefficient of thermal ionization.
β = Coefficient of recombination.
F = Factor depending upon the probability of ionization.
σ = Radius of cross-section of collision between neutral molecules, cm.
J = Mechanical equivalent of heat, joule/cal.
C_v = Specific heat of gas at constant volume, cal/g/°C.
t = Time, sec.
S = Heat radiated from the surface of the core, cal/cm^2/T^4.

Correspondence on Monographs is invited for consideration with a view to publication.
Dr. Yussuf and Prof. Prescott are at King's College, Newcastle upon Tyne.

(2) INTRODUCTION

It has been noticed that arc and glow oscillators are difficult to operate at high frequencies, since the arc path then loses the negative resistance necessary for oscillation.[6,7] Also the frequency of oscillation differs from that of the tank circuit as if the arc path is inductive. The presence of an inductive effect in the ionized path can be verified using a cathode-ray oscillograph. If an alternating current of small amplitude is superimposed on a direct-current arc and the alternating voltage developed between the electrodes is applied to the Y-plates and a voltage proportional to the alternating component of the current to the X-plates, an ellipse will appear on the screen. The shape and the slope of the ellipse will change with the frequency of the injected current.

The behaviour of the discharge path may be due to the characteristics of either the positive column or the cathode or of both. Although the behaviour of the positive column has been explained qualitatively to be due to the time lag of ionization,[1] no attempt at a thorough investigation was made. The authors know of only one attempt to investigate the high-pressure positive column of an arc discharge in mercury vapour. This was made by Von Weizel, Rompe and Schulz[2] in an attempt to explain the experimental results of Magnold[8] and was not conclusive. In this paper, the behaviour of the positive column of the mercury-arc discharge at high and at low pressures is considered separately. The high-pressure case is considered first. A theoretical expression for the impedance of the column as a function of frequency is developed. This is then compared with the measurements given by Magnold. The low-pressure case is then considered. Again a theoretical expression for the impedance is developed. In addition, measurements of the impedance were made for a range of conditions and the results are compared with the theory.

The behaviour of the cathode will depend upon the kind of emission obtained from its surface. In 1930 Buchtiger[7] investigated the behaviour of the hot cathode (where thermionic emission occurs) in an arc discharge of very short length, when an alternating current was superimposed on the main direct current. Buchtiger's work was based on a voltage/current static characteristic of the d.c. arc obtained experimentally, and it was then shown that owing to the large thermal capacity of the metal cathode, its inductive effect would be negligible if the operating frequency were above a certain value, depending upon the dimensions and material of the cathode. At higher frequencies the cathode region behaves as a path of constant positive resistance independent of frequency. The same results were obtained by the authors from the starting point of the static voltage/current characteristic derived by Druvestyn and Penning[9] for a hot cathode arc of very short length.

No similar work has been done on the high field-emission type of cathode. Actually no acceptable theory has yet been developed for the phenomena at its cold surface. In the arc discharge in mercury vapour at low pressure, a relatively high voltage appears across the cathode region. A basic assumption of the experimental method adopted in this work is that the behaviour of the cathode region will not materially affect the impedance measured

between the electrodes. In order to justify this assumption, the positive column under consideration was of considerable length.

(3) GENERAL PROPERTIES OF THE PLASMA

The term "plasma" is applied to an ionized-gas medium in which the concentration of electrons and positive ions are approximately equal and relatively high.[3] Such a plasma is usually established by an electric field. In a d.c. plasma, the electrons, the positive ions and the neutral molecules are in thermal equilibrium but may not be at the same temperature.[16] The charge carriers in moving along the electric field pick up energy from the field and deliver it to the neutral gas molecules during the collisions which take place between them. This energy, delivered to the gas, escapes by radiation, conduction and convection.

The electrons and the positive ions are continuously lost from the plasma by their volume recombination and their diffusion outwards if there is any charge-carrier concentration gradient. In order to sustain the plasma it is therefore necessary to produce by ionization as many electrons and positive ions as are lost by recombination and diffusion. The mechanism of the ionization process and the way in which the electrons are lost from the volume, together with the temperature of the charged and neutral particles, depend mainly on the gas pressure.[16]

When the pressure is low, say 0·1 cm Hg or less, the electron mean free path is sufficiently long for the electron to absorb a considerable amount of energy from the electric field between each two successive collisions with the neutral gas molecule. The result will be a high electron temperature (2×10^4 °K) while the gas temperature[3] is as low as 30° C. The electrons will then have sufficient energy to ionize the neutral molecules when they collide. In such cases electron ionization is the main process of electron production. The rate of production of electrons per cubic centimetre of the volume will be αN, the value of α being a function of the electron temperature.[4,14] Again, with high electron temperatures, direct recombination in the volume between an electron and a positive ion will be unlikely.[3] They will diffuse to the walls of the container, where they deliver their thermal energy and recombine there, releasing the ionization energy.

However, in a high-pressure plasma (0·5 atm or more), the electron mean free path is short, and the temperatures of the electron, the positive ion and the gas are nearly the same (5 to 6×10^3 °K). In such cases the ionization process will be purely thermal,[10] and volume recombination accounts for the loss of electrons from the plasma. As thermal ionization depends on the number of collisions between the neutral molecules, the rate of production of electrons per cubic centimetre will be $\alpha_1 N_g^2$, the coefficient α_1 depending upon the gas temperature.

(3.1) Drift Current through a Plasma

In a plasma sustained by a d.c. electric field, the electrons, the positive ions and the neutral molecules move in random directions. Each kind of particle will have its distribution of velocities, which is usually Maxwellian. However, the electrons being negatively charged will drift with an average velocity or mobility against the direction of the field, and the positive ions will drift in the direction of the field. The total current drawn from the plasma will be the sum of the two drift currents. The positive-ion current is very small and can be neglected compared with that of the electrons, owing to the smaller mass and higher velocities of the electrons. Considering a cylinder of 1 cm^3 volume and 1 cm^2 cross-section, placed coaxial with the electric field, the current density of the plasma along the axis will be

$$\iota = \mu e N \mathscr{E}$$

Now any change, $\Delta\iota$, in the drift current will be produced corresponding changes ΔN and $\Delta\mathscr{E}$ or

$$\Delta\iota = \mu e \mathscr{E}(\Delta N) + \mu e N(\Delta\mathscr{E})$$

The change ΔN can be produced in one or more of the followi ways.

(*a*) Electrons pouring into the cylinder in the direction of the ax In arc discharges, this will correspond to increased emission fr the cathode and will not be considered here.

(*b*) Changes in the radial flow of electrons and ions, i.e. chang in the radial diffusion.

(*c*) In this case there are no changes either in the radial or t axial flow of electrons. As a result of the changes in the number collisions between particles already present in the volume co sidered, new electrons are formed. It is clear that this will be d to changes in ionization.

In order to show in a clear and simple way the behaviour diffusion and ionization and their effect on the electron co centration when these changes are sinusoidal, and to reach understanding of their physical meaning, each effect will considered separately.

(3.1.1) Effect of Diffusion on Electron Concentration.

Let it be assumed that it is possible to have a cylinder of ur form concentration placed coaxially with a container of relative large radius. If a sinusoidal quantity ($\Delta\xi e^{j\omega t}$) of electro be introduced into the core, some of these electrons w remain in the core raising the concentration and the rest w diffuse outwards from the boundary surface of the core. Th sinusoidal changes due to diffusion alone in the electron co centration at points outside the core along the radius will defined[3] by

$$\frac{d^2n}{dr^2} + \frac{1}{r}\frac{dn}{dr} = \frac{1}{D}\frac{dn}{dt} = \frac{j\omega}{D}n$$

If this differential equation is solved with the boundary co dition $n = n_1$ at $r = a$ and the concentration is finite at a ve large radius, it is shown in Section 6.1 that the amplitude, Δ of the number of electrons which alternates outwards a inwards at the boundary surface $r = a$ will be

$$\Delta q = \pi D n_1 [1 + 2a\sqrt{(j\omega/D)}]/j\omega$$

and

$$\Delta q/\Delta\xi = D[1 + 2a\sqrt{(j\omega/D)}]\Big/ j\omega\left\{a^2 + \frac{D}{j\omega}[1 + 2a\sqrt{(j\omega/D)}]\right\}$$

Fig. 1 shows the variation of $\Delta q/\Delta\xi$ against frequency f different values of the coefficient of diffusion D. The shape

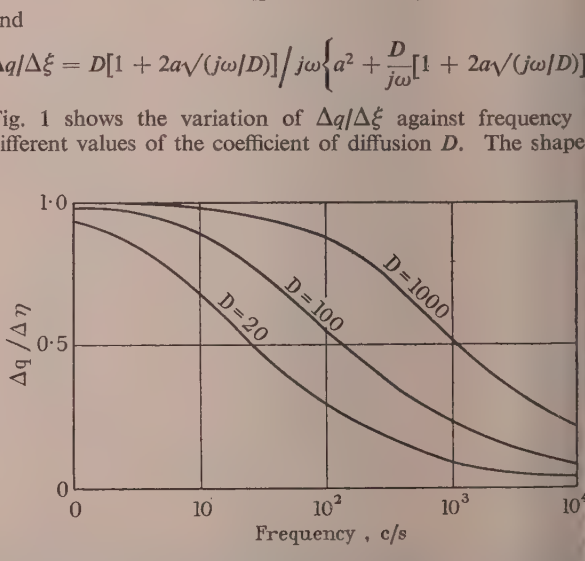


Fig. 1.—Curves illustrating the change with frequency in the ratio the number of diffusing electrons to the number introduced i the volume.

curves will not be difficult to explain if it is remembered that electron needs a certain time to diffuse from the core. When frequency is low and the periodic time is long, $\Delta q/\Delta\xi = I$, the electrons introduced in the core will diffuse and will be hanged between the core and its surroundings. However, as frequency is increased the periodic time is shorter and only a ction of the electrons introduced will be able to get out of the e. At very high frequencies, practically no electrons will take t in the exchange, and the introduced electrons will be utilized raise the concentration in the core, which will then be $= \Delta\xi/\pi a^2$ as $\omega \to \infty$.

The effect of the coefficient of diffusion D is clear. For large ues of D, i.e. higher electron velocities, the diffusing electrons follow the rapid changes more readily, and diffusion con- es to occur at higher frequencies than when D is small.

From the above it is seen that the effect of diffusion on the tron concentration can safely be neglected in the high- ssure plasma as $D \simeq 3$, and the order of frequency which be considered is above 500 c/s.

.2) Effect of Ionization on Electron Concentration.

n order to show the effect of ionization alone, let a cubic timetre, of uniform concentration N, be considered. In the e of a low-pressure plasma the ionization process will be per- ned by electrons. If there is no loss of charge from the ume either by diffusion or recombination, the electron con- tration will continue to rise at a rate proportional to N or $/dt = \alpha N$. If now a number of electrons varying sinusoid- , $ne^{j\omega t}$, is extracted from the volume, the original rate of of concentration will be reduced. In order to combat this uction an increase, $\Delta\alpha e^{j\omega t}$, in the value of α would be neces- . The original rate of rise of concentration would be restored ne relation

$$(N - n)(\alpha + \Delta\alpha) - \frac{dn}{dt} = \frac{dN}{dt} = \alpha N$$

atisfied. As the changes are sinusoidal, then

$$(N - n)(\alpha + \Delta\alpha) - j\omega n = \alpha N$$

glecting the term of second order we obtain

$$n = \frac{N\Delta\alpha}{\alpha + j\omega}$$

m the above expression it is seen that the number of electrons, roduced by ionization and available for extraction from the ume, diminishes with increased frequency. Furthermore, it behind the change $\Delta\alpha$, the lag increasing with frequency.

A similar treatment indicates the same physical behaviour in case of a high-pressure plasma when the ionization is thermal. his case if there is no recombination or diffusion, the rate of of electron concentration will be $dN/dt = \alpha_1 N_g^2$. If now α_1 ncreased by a small value $\Delta\alpha_1$, the change in the rate of rise be

$$\Delta(\alpha_1 N_g^2) = 2\alpha_1 N_g(\Delta N_g) + N_g^2(\Delta\alpha_1)$$

restore the rate of rise to its original value, a quantity ΔN of trons should be extracted from the volume such that

$$\Delta(\alpha_1 N_g^2) - \frac{d}{dt}(\Delta N) = 0$$

he changes are sinusoidal

$$2\alpha_1 N_g(\Delta N_g) + N_g^2(\Delta\alpha_1) - j\omega(\Delta N) = 0$$

But as $N_g + N = N_0$, the normal concentration of the gas, ΔN_g, equals $-\Delta N$, whence

$$\Delta N = \frac{N_g^2(\Delta\alpha_1)}{2\alpha_1 N_g + j\omega}$$

From this it is seen that whether the ionization is performed by electrons or by neutral molecules, its effect on the electron concentration will depend on the frequency. At low frequencies, the periodic time is long enough for all the capable particles to take part in the ionization process. However, as the frequency is increased, only the higher-speed particles will be able to do so. At very high frequencies the changes are so rapid that very few of the particles will be able to take part in the ionization process and $\Delta N = 0$; thus no changes in electron concentration will be produced.

(4) APPLICATION TO ARC DISCHARGES

A notable example of the plasma is the positive column of the arc discharge. However, a part of the voltage appearing between the electrodes of the arc will be developed across the cathode region. The changes in the current due to any change in the arc voltage, and therefore the effective impedance of the path, will be governed by the behaviour of either the positive column at the cathode or of both.

The cathode of an arc discharge may be either of the high-field-emission type or the hot-cathode type when thermionic emission occurs, depending upon the material of the cathode.[6,3] The behaviour of the hot cathode when an alternating current of small amplitude was superimposed on the d.c. arc of short length has been investigated by Buchtiger[7] and recently by the authors. It was found that the effective impedance of such an arc will be inductive at very low frequencies (1 c/s) and changes to a constant positive resistance at higher frequencies (above 80 c/s) depending upon the material and dimensions of the cathode. No similar work has been done on the field-emission-type cathode. However, the behaviour of the cold cathode can be masked to a certain extent if the length of positive column is increased considerably.

Since there are large differences in character and behaviour between the low-pressure and the high-pressure positive columns, each will be considered separately.

(4.1) Positive Column of a High-Pressure D.C. Arc in Mercury Vapour

(4.1.1) General.

The positive column is a region of uniform but relatively high potential gradient, the value of which depends on the arc current and nature of the gas. In regions far away from the electrodes, where their cooling effect is small, the gas temperature in the column is high, of the order of 5 to 6×10^3 °K. With the pressure encountered the mean free path is short, and the electrons and positive ions will have practically the same temperature[16] as that of the gas molecules ($T = T_g$).

Besides the neutral gas molecules, each unit volume in the column contains an equal number of electrons and ions. The electron concentration is usually high, about 10^{14} per cubic centimetre. In steady states, the electron concentration per cubic centimetre of the column was found by Saha to be $N^2/N_g = AT^{3/2}\epsilon^{-eV_i/kT}$, where A is a constant. Although the Saha expression was originally derived for the thermal ionization in the solar atmosphere and based on thermodynamical reasoning, yet he pointed out its possible application in arcs.

The validity of the Saha expression in high-pressure arc discharges was justified in the lengthy and careful analysis of Ornstien.[10] Considering a Maxwellian distribution of particle

velocities, the process of ionization in the column was shown to be a thermal and reversible process. Two high-speed neutral molecules, with enough energy between them, can produce an electron and a positive ion when colliding with each other. The rate of ionization will then be $\alpha_1 N_g^2$. The completely reversible process of recombination will be a double collision. The electron and positive ion when colliding with each other will form a neutral molecule with a surplus energy equal to the ionization energy. Such a molecule must then collide with another neutral molecule to get rid of its surplus energy. The rate of recombination will then be $\beta N_g N^2$. The rate of increase of electron concentration in the column will be given by

$$\frac{dN}{dt} = \alpha_1 N_g^2 - \beta N_g N^2$$

For a d.c. arc, $dN/dt = 0$ and

$$\alpha_1/\beta = N^2/N_g = AT^{3/2}\epsilon^{-eV_i/kT} \quad . \quad . \quad . \quad (5)$$

The coefficient of thermal ionization, α_1, can be determined from the number of collisions between neutral molecules of the proper velocities. It was given by Ornstien to be

$$\alpha_1 = 8F\sigma^2\sqrt{(\pi kT/M)}\epsilon^{-eV_i/kT} \quad . \quad . \quad . \quad . \quad (6)$$

The coefficient of recombination β can be obtained from eqns. (5) and (6).

(4.1.2) Temperature and Radius of the Positive Column.

The electrons and positive ions absorb a certain amount of energy from the source sustaining the discharge as they drift axially in the electric field. The power input to the column balances the various losses.[3,9] These losses are mainly due to the diffusion of the charged particles from the column outwards, carrying their thermal and ionization energies, and to the heat lost by the hot gas by convection, conduction and radiation. The cooling effect of the convection currents can be eliminated if the tube containing the arc is placed horizontally.

The proportion of the heat conducted depends upon the kind of gas. In certain gases, the radiation spectrum from the column consists of bands of wavelengths separated by regions of no radiation. In such cases the heat radiated is small while that conducted is fairly large. However, for an arc in mercury vapour it was found by Elinbass[11] that the radiation spectrum is almost uniform, and the heat radiation is over 70% of the input.

Owing to the continuous cooling at the boundary of the column, the gas temperature at points along the radius of the container will not be uniform. This radial temperature variation will cause a radial change in electron concentration. However, the positive column has been treated in general as a core of uniform concentration, and of an effective radius and temperature depending on the value of the arc current. The effective temperature and radius of the core have been determined experimentally.[12,13] An approximate method of determining these values for an arc in mercury vapour will now be given, based upon the following assumptions:

(*a*) The core of the arc burns axially in a horizontal container to eliminate the effect of the convection currents.

(*b*) The total heat radiated from the positive column will be considered as radiated from the core. Actually there is a certain amount of heat radiated from the gas between the core and the walls of the container. This radiation will be neglected since it depends on T^4 and the temperature decreases rapidly.

Consider a cylinder of radius r, and thickness dr; then the amount of heat conducted per second is

$$Q = -K2\pi r dT/dr \quad . \quad . \quad . \quad . \quad . \quad (7)$$

The coefficient of thermal conductivity, K, in mercury vapo was given by Elinbass as

$$K = 2{\cdot}1 \times 10^{-7}T^{3/4} \text{ thermal mho/cm}$$

Substituting K in eqn. (7) and integrating

$$\log r = -\frac{7{\cdot}536}{Q} \times 10^{-7}T^{7/4} + C \quad . \quad . \quad .$$

The constant C can be determined if the radius and temperatu of the walls of the container are known.

If we consider the total arc current to be carried by electro only, then in the case of a uniform concentration in the core,

$$I = \pi r^2 \mu e N \mathscr{E}$$

But the electron concentration as given by the Saha expressio

$$N^2 = N_g A T^{3/2}\epsilon^{-eV_i/kT}$$

Therefore $$r = \sqrt{\left[\frac{I}{\pi\mu e E\sqrt{(AN_g T^{3/2}\epsilon^{-eV_i/kT})}}\right]} \quad . \quad . \quad .$$

The point of intersection of the two T/r curves defined eqns. (8) and (9) will give the effective radius and temperatu of the positive column. The two curves were calculated fo column placed in a tube of 1 cm radius under the followi conditions, which were the same as those of Elinbass: $I = 5$ an $\mathscr{E} = 8$ volts/cm, $Q = 10$ watts, and the temperature of walls atmospheric pressure was 900° C. Fig. 2 shows the approxima

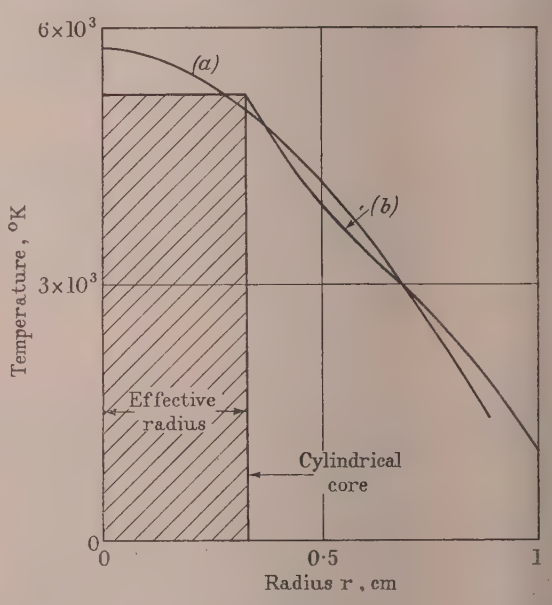


Fig. 2.—Temperature change along the radius of the tube contai
(*a*) Exact by Elinbass.
(*b*) Approximate.

temperature distribution along the radius, together with the ex one given by Elinbass. The assumptions previously made do seem to have altered the final results to any great extent.

(4.1.3) Impedance of the Positive Column to Alternating Currents.

Before the mathematical analysis, it is useful to consider expected behaviour of the column when an alternating curren superimposed on the direct current. If the arc current is m to vary sinusoidally at a very small rate, the gas temperatu

e electron concentration and the electric field would be ex-cted to undergo similar sinusoidal changes. During the cycle, e positive column would then move from one state of thermal uilibrium to another, obeying the Saha and Maxwell laws. henever it passed the temperature T, it would always have the ctron concentration given by Saha and the same number of ctrons at a specified energy level as given by Maxwell.

As the ionization in the column is purely thermal the electrons ll play only an indirect part in the ionization process. When e electric field is rising above its d.c. value, the electrons are celerated and transfer their energy to the gas molecules without oducing ionization. The average temperature of the gas will erefore increase, resulting in an increase of the thermal ioniza-n and the electron concentration. When the electric field ls below its mean value, recombination takes place, producing olecules with surplus energy which will collide with other olecules to lose that energy. The ionization cycle would then reversible.

This state of affairs can only be maintained if the rate of ange with time is so slow that all the molecules capable of ing so take part in the ionization process. However, if the quency is increased, the rate of change of electron concentra-n will be higher than the average rate of production of ctrons. The slower molecules will fail to transfer their energy ionize other molecules, and only the faster molecules will be le to do so. As the frequency is increased further, fewer olecules will take part in the ionization process. At very high quencies the number of ionizing collisions are too small for y change of electron concentration to occur.

The impedance per unit length of a conducting path is $\Delta\mathscr{E}/\Delta I$. evaluate this expression in the dynamic case, for the positive lumn of an arc in mercury vapour, the following assumptions e made:

(*a*) The positive column can be represented by a cylindrical core of uniform electron concentration and temperature placed axially in a horizontal tube to eliminate the cooling effect of convection currents.

(*b*) The ionization process is a completely reversible one, as shown by Ornstien, and therefore

$$dN/dt = \alpha_1 N_g^2 - \beta N_g N^2$$
$$= \beta N_g\left(\frac{\alpha_1}{\beta}N_g - N^2\right) \quad . \quad . \quad . \quad . \quad (10)$$

At the steady state

$$dN/dt = 0$$
$$\alpha_1/\beta = N^2/N_g = AT^{3/2}\epsilon^{-eV_i/kT} = f(T) \quad . \quad . \quad . \quad (5)$$

(*c*) The frequency of operation is so high that the effect of diffusion can be neglected.

(*d*) The frequency of operation is so high that the exchange of heat by conduction between the column and its surroundings in the dynamic case is negligible. This assumption is nearly true since the differential equation for heat conduction is the same as that for diffusion, and the coefficient of thermal conductivity of gases is small.

If the vapour temperature is made to vary sinusoidally by $Te^{j\omega t}$, the electron concentration will undergo similar changes $Ne^{j\omega t}$. A relation between ΔN and ΔT can be obtained by ferentiating eqns. (10) and (5) and substituting the steady-state ation

$$\Delta N = N_g\Delta(\alpha_1/B)/[(\alpha_1/B) + 2N + j\omega/(BN_g)]$$
$$= \frac{N_g f'(T)}{\frac{\alpha_1}{B} + 2N + j\omega/(BN)}\Delta T \quad . \quad . \quad . \quad . \quad . \quad (11)$$

e electron velocities in the column are much higher than those the positive ions as they have a smaller mass. The total arc current can be assumed to be carried by electrons only and to be given by

$$I = \pi a^2 N\mu e\mathscr{E}$$

Introducing the small alternating changes, $\Delta Ie^{j\omega t}$ and $\Delta\mathscr{E}e^{j\omega t}$, to the current and electric field, respectively

$$\Delta I = \frac{I}{N}\Delta N + \frac{I}{\mathscr{E}}\Delta\mathscr{E}$$

or
$$\Delta N = \frac{N}{I}\left(\Delta I - \frac{I}{\mathscr{E}}\Delta\mathscr{E}\right) \quad . \quad . \quad . \quad . \quad (12)$$

For conservation of energy, the change, $\Delta\mathscr{E}I$, in the power input per centimetre to the column should be equal to the change in the various losses. These are as follows:

(*a*) Heat should be supplied to the gas to raise its temperature by an amount ΔT. The rate of supply of electrical energy will be $j\omega JmgC_V\Delta T = j\omega h\Delta T$.

(*b*) To increase the electron concentration by ΔN the gas molecules must absorb from the supply the necessary ionization energy. The rate of supply of electrical energy will be $j\omega\pi a^2 eV_i\Delta N = j\omega gN\Delta$.

(*c*) If the amount of heat conducted to the outer layers surrounding the core is neglected in the dynamic case, the rest of the power input to the column will be radiated. In the steady state, the heat radiated per second per centimetre length of a column at a temperature T is $J2\pi aST^4$. The heat radiated in the dynamic case will be $8J\pi aST^3\Delta T = C\Delta T$.

Summing up, $\Delta(\mathscr{E}I) = \mathscr{E}\Delta I + I\Delta\mathscr{E}$
$$= (C + j\omega h)\Delta T + j\omega g\Delta N \quad . \quad . \quad . \quad (13)$$

The variables ΔT and ΔN can be eliminated between eqns. (11), (12) and (13). Rearranging,

$$\frac{\Delta\mathscr{E}}{\Delta I} = -\frac{\mathscr{E}}{I}\cdot\frac{1 - \frac{N}{\mathscr{E}I}\left\{j\omega g + \frac{(C + j\omega h)[(\alpha_1/\beta) + 2N + j\omega/(BN_g)]}{N_g f'(T)}\right\}}{1 + \frac{N}{\mathscr{E}I}\left\{j\omega g + \frac{(C + j\omega h)[(\alpha_1/\beta) + 2N + j\omega/(BN_g)]}{N_g f'(T)}\right\}}$$

The above expression for the impedance can be split into a resistive and a reactive component. If the latter is divided by $j\omega$ it will give an effective inductance per centimetre of the column. The two components are calculated and plotted against frequency in Fig. 3 for a positive column having the following values:

Gas pressure	Atmospheric
Length of column	3 cm
Direct arc current	5 amp
Longitudinal electric field, $\mathscr{E}$	8 volts/cm
Effective radius, a	0·33 cm
Effective temperature	5 250° K
Heat radiated in the d.c. case	30 watts

In Fig. 4 the calculated resistance/frequency curve is drawn, together with the experimental one obtained by Magnold, on the same frequency scale.

(4.2) Positive Column of the Low-Pressure D.C. Arc in Mercury Vapour

(4.2.1) General.

The positive column of the low-pressure arc occupies nearly all the space between the electrodes except a narrow region near the cathode. It is a region of almost uniform longitudinal voltage gradient of a magnitude which is dependent upon the arc current and the vapour pressure.[3] Besides the neutral vapour molecules,

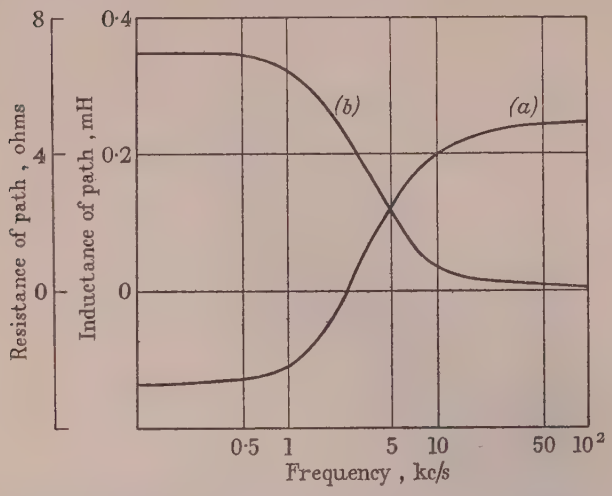


Fig. 3.—Calculated resistance and inductance of a 3cm 5-amp d.c. positive column at atmospheric pressure.

(*a*) Resistance.
(*b*) Inductance.

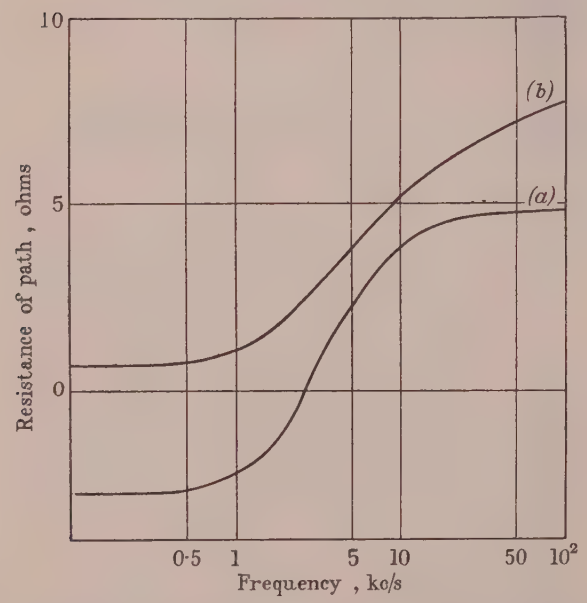


Fig. 4.—Calculated resistance of a 3cm 5-amp positive column at atmospheric pressure, compared with the experimental results obtained by Magnold.

(*a*) Calculated.
(*b*) Experimental by Magnold.

each cubic centimetre of the positive column contains an equal number of electrons and positive ions moving at random in all directions. However, the electrons being negatively charged drift with an average velocity or mobility towards the anode, while the positive ions drift towards the cathode. At the walls of the tube containing the arc, electrons and ions continuously recombine. Such recombination sets up a charged-particle concentration gradient along the radius of the tube. Electrons and ions will diffuse from the column to the walls under the effect of this gradient. The number[3] of charged particles per square centimetre per second diffusing in a direction normal to the flow will be $-DdN/dr$ at any radius r. The total number of ions and electrons diffusing to the wall per centimetre per sec will be $-2\pi aD(dN/dr)_{r=a}$. Besides diffusion, the loss of i and electrons in the column may be due to their direct recombi tion when they collide with each other. However, in low-press arcs the effect of volume recombination is small and can neglected.

For the positive column to be sustained, electrons and i have to be produced in the column at the same rate as they lost by diffusion. In low-pressure arcs the process of ionizat is performed mainly by electron collisions with the neu vapour molecules. Considering the electrons to have a M wellian distribution of velocities, and ionization to occur a single impact, Killian[4] gave the rate of ionization per second electron in mercury vapour as

$$\alpha = 6{\cdot}24 \times 0{\cdot}28 \times 10^5 \frac{P}{T_g}\sqrt{(T)}\left(V_i + \frac{2kT}{e}\right)\epsilon^{-eV_i/kT}$$

α and D determine the electron concentration at different poi along the radius of the tube container. In cases when the vap pressure is high enough to neglect Langmuir[5] space charges the walls, Schottky[1] defined the concentration of electrons cubic centimetre by the equation

$$\frac{d^2N}{dr^2} + \frac{1}{r}\frac{dN}{dr} + \frac{\alpha}{D}N = \frac{1}{D}\frac{dN}{dt}$$

For a d.c. column $dN/dt = 0$, and the solution will be gi by the Bessel function $N = N_0 J_0 r\sqrt{(\alpha/D)}$. As the containe of insulating material, the concentration falls to zero at walls where $r = a$. The total number of electrons per centime length of the column will be $\eta = \int_0^a 2\pi rN dr$. The charg particles in the column drift towards the electrodes carrying arc current I in the electric field $\mathscr{E}$. The power input $\mathscr{E}I$ centimetre of the column is dissipated in two forms.[4] A part is transferred from the charged particles to the neutral molecu in the various collisions that take place between them. Most this power is transferred by the electrons as they make a grea number of collisions with the neutral molecules than do positive ions. These collisions are either elastic, when electron loses a small fraction of its energy at impact, or inelast when it loses a fixed amount of energy depending on the kind excitation produced. If, on the average, the electron loses fraction f of its energy at each impact, and considering a M wellian distribution of electron velocities, the rate of transfer energy from the electrons to the vapour molecules will be[3]

$$P_1 = \frac{2mf}{\lambda\sqrt{\pi}}v_0^3\eta = H\eta T^{3/2} \quad \text{where} \quad H = \frac{2f(2K)^{3/2}}{\lambda\sqrt{(\pi m)}}$$

Most of this power is conducted through the gas to the walls the container and the rest is radiated.

The remaining part P_2 of the input power is delivered by charged particles to the walls when they reach it, as a result diffusion and recombination releasing their ionization energy,

$$P_2 = \alpha\eta(V_i + kT + V_p)$$

The value of P_2 is usually less than 7–8% of the total input po except in cases when the vapour pressure is very low. T

$$P = P_1 + P_2$$

(4.2.2) Effective Impedance.

The impedance of a conducting path is the rate of chang the voltage divided by the current. For a unit length of column, the impedance will be $\Delta\mathscr{E}/\Delta I$. Before any attemp

ade to derive this impedance let us consider how the current the column changes with alterations in the electric field. creasing the electric field will increase the electron temperature, the electrons will pick up more energy in between each two ccessive collisions with the neutral vapour molecules. This ll result in a higher rate of production of ions. Some of the ctrons so produced will diffuse to the walls of the container d the rest will remain in the column to be collected by the ctrodes. The change in the current depends mainly on the ctron temperature and on the number of electrons per centi- tre of the column.

If the above changes are sinusoidal, the time element has to taken into account. For low frequencies the periodic time is ng enough for all the capable electrons to transfer the energy ey gained from the varying field to the neutral molecules and nize them. Again, the increasing number of electrons will have fficient time to diffuse from the column outwards to the walls. owever, as the frequency is increased and the periodic time is duced, the above changes will be incomplete. The number of ctrons taking part in the ionization process will be reduced, ce it will be limited to the higher-speed electrons. Similarly, e number of diffusing electrons will be less if the periodic time short. At very high frequencies the electron concentration will t change, and it will remain constant, independent of the pidly varying electric field.

In order to put the above picture in mathematical form, it will necessary to determine the alternating changes in electron ncentration for a corresponding change in the electric field. e radial electron concentration was defined by Schottky to be

$$\frac{d^2N}{dr^2}+\frac{1}{r}\frac{dN}{dr}+\frac{\alpha}{D}N=\frac{1}{D}\frac{dN}{dt}$$

the d.c. column $dN/dt = 0$ and $N = N_0 I_0 r\sqrt{(\alpha/D)}$. Assuming coefficient of diffusion D to remain constant for small ernating changes of electron temperature, the alternating ncentration $ne^{j\omega t}$ is shown in Section 6 to be defined by

$$\frac{d^2n}{dr^2}+\frac{1}{r}\frac{dn}{dr}+\frac{\alpha-j\omega}{D}n+\frac{\Delta\alpha}{D}N=0 \quad . \quad . \quad (14)$$

the walls of the container there will be no changes of electron ncentration, the wall being an insulator, i.e. $(n=0)_{r=a}$. ving the differential eqn. (14), we obtain

$$n=\frac{\Delta\alpha}{j\omega}N_0 J_0 r\sqrt{(\alpha/D)}$$

e change in the total number of electrons per centimetre length l be

$$\Delta\eta=\int_0^a 2\pi rndr=\frac{\Delta\alpha}{j\omega}\eta$$

$$=\frac{\eta(d\alpha/dT)}{j\omega}\Delta T \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

Considering the total current to be carried by electrons only,

$$I=\mu e\eta\mathscr{E}$$

l for small sinusoidal changes

$$\Delta I=\frac{I}{\mathscr{E}}\Delta\mathscr{E}+\frac{I}{\eta}\Delta\eta \quad . \quad . \quad . \quad . \quad . \quad (16)$$

nsidering the power input $\mathscr{E}I$, per centimetre of the column, part P_2 (about 7–8%) given to the walls can be neglected, ecially when the periodic time is too short for diffusion to occur. The changes in the power input will be due to the changes in the power delivered to the gas, or

$$\Delta(\mathscr{E}I)=\Delta(P_1)=\Delta(HT^{3/2}\eta)$$

$$=\frac{3\mathscr{E}I}{2T}\Delta T+\frac{\mathscr{E}I}{\eta}\Delta\eta \quad . \quad . \quad . \quad (17)$$

The value of $\Delta\eta$ was determined from eqn. (15). Eliminating $\Delta\eta$ and ΔT between eqns. (15), (16) and (17), rearranging and putting $A = 3/2T(d\alpha/dT)$

$$\frac{\Delta\mathscr{E}}{\Delta I}=\frac{\mathscr{E}}{I}\frac{j\omega A}{2+j\omega A}=\frac{\mathscr{E}}{I}\left(\frac{\omega^2A^2}{4+\omega^2A^2}+j\omega\frac{2A}{4+\omega^2A^2}\right)=R+j\omega L$$

(4.2.3) **Experimental Investigations.**

The experimental work carried out was intended to determine and investigate the following:

(*a*) Measurement of the electron temperature and electric field in the column of a certain direct arc current. These values are used to calculate the derived expression for the dynamic impedance of a column of known length.

(*b*) To investigate the effect of the length of the column on the impedance.

(*c*) To determine the behaviour of the impedance with frequency for different values of the direct arc current and vapour pressure.

The discharge apparatus shown in Fig. 5 was specially designed. It consisted of two 2 in-diameter tubes, 150 cm and 75 cm long,

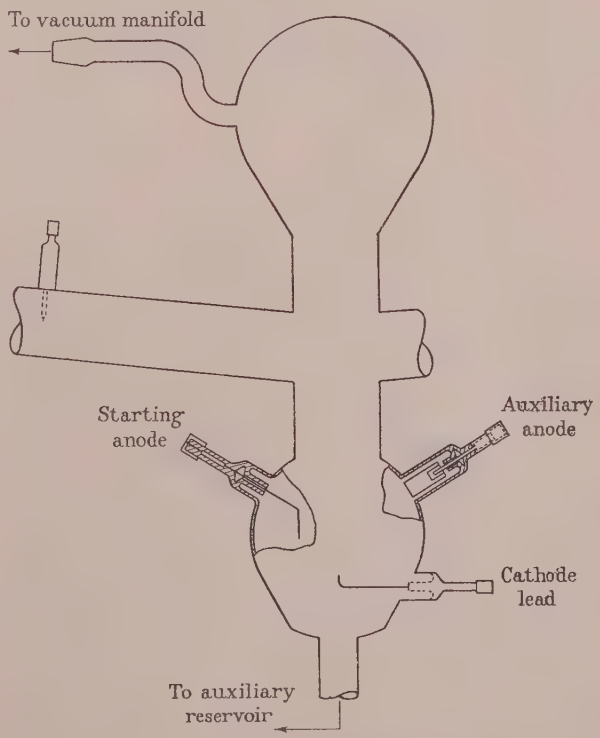


Fig. 5.—Discharge apparatus.

respectively. Both tubes had a common mercury pool cathode. Two molybdenum wire probes, as shown in Fig. 6, each of 0·0314 cm^2 collecting area, were fixed in each tube for determination of electron temperature, concentration and electric field along the axis of the tube. In addition, a movable probe (see Fig. 7), fitted with a piece of soft iron and an electromagnet was provided in the 150 cm tube so as to give the electron concentration

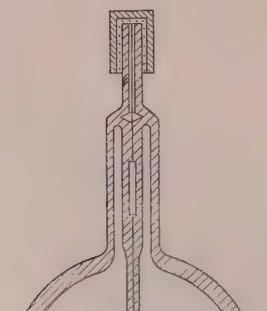


Fig. 6.—Fixed probe.

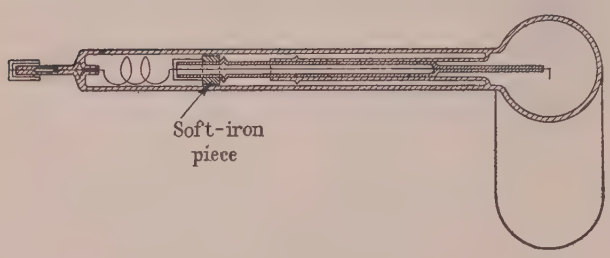


Fig. 7.—Movable probe in the 150cm tube.

and radial electric field at different points along the diameter. The probes were 8 in and 12 in apart in the short and long tube respectively. The temperature of the walls was measured using a calibrated nickel wire wound round the tube. The vapour temperature was taken to be that of the walls. An electric mat heater was wound round each tube to prevent the condensation of the vapour on the walls when the cathode-pool temperature was higher than that of the room.

The lower bulb containing the mercury pool was placed in a water bath and kept at the required temperature using a thermostat control. The bulb was provided with three electrodes. The lower one immersed in mercury was used as a cathode lead. One of the top electrodes was used as an auxiliary anode, and the other for starting. Starting was effected by changing the level of the mercury in the pool to make and break the starting anode circuit, using an auxiliary mercury reservoir. The whole apparatus was continuously evacuated by a mercury diffusion pump backed with a rotary oil pump.

Before any test was carried out, the lower bulb was partially filled with mercury and the apparatus evacuated. A heavy discharge current was then passed in each tube to ensure complete degassing of the electrodes. This was carried out on five runs of eight hours each. In between the tests which followed, the apparatus was always under vacuum and no air was allowed to enter. Two groups of tests were made; the first at a vapour pressure corresponding to a bath temperature of 20° C, and the second at 40° C. The group contained a series of tests carried on each of three different direct arc currents, namely 1·5, 2·5 and 4 amp. The discharge current was passed through one of the two tubes, and the electron temperature, the electric field and the effective impedance at different frequencies were measured for each current. The tests were then repeated when the current was passing in the second tube. Each separate type of measurement is discussed below.

(4.2.4) Determination of Electron Temperature and Electric Field.

Considering each tube separately, the arc current was adjusted to the required value and kept constant during the run of the measurements.

Readings of the current collected by each of the fixed probes were taken for different values of its potential relative to that of the cathode, the other probes being left unconnected. The readings were used to determine the electron concentration, temperature and the longitudinal electric field along the axis of the tube. For the complete theory of probe measurements reference is made to the works of Killian and Langmuir.[4,5] The characteristics of the two fixed probes in the short tube are given in Fig. 8 for an arc current of 1·5 amp at 40° C bath temperature.

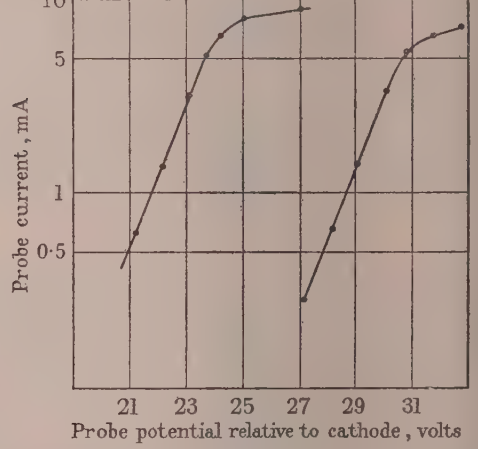


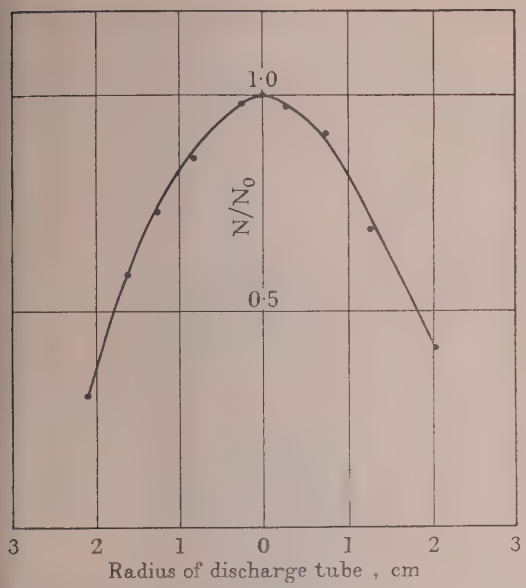
Fig. 8.—Characteristics of the two fixed probes in the 75cm tube.

Arc current	1·5 amp
Bath temperature	40° C
Distance apart of the two probes	8 in

The voltage/current characteristics of the movable probe were then taken for different positions of the probe along the tube diameter. The electron concentration and radial electric field at different points along the diameter of the 150 cm tube are given in Figs. 9 and 10 respectively for a 1·5-amp arc current and 40° C bath temperature. The experimental results agree well with the Schottky equation and with the results of both Killian and Langmuir.

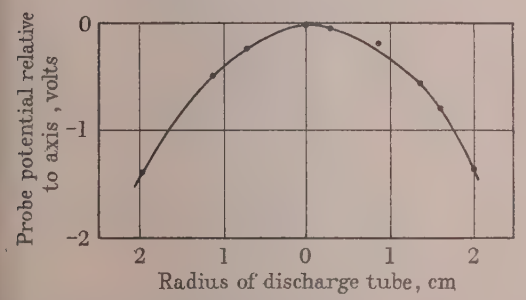
(4.2.5) Measurement of the Effective Impedance.

A small alternating current of about 100 mA and of adjustable frequency was injected into the arc path from a power amplifier. The impedance of the arc path was then measured on the bridge, shown in Fig. 11, for different frequencies of the injected current. A choke coil was placed in series with the d.c. source to eliminate its shunting effect. When the bridge is balanced the effective resistance of the arm containing the arc is $R = C_1R_2$ and its inductance is $L = C_2R_1R_2$. This bridge was convenient since the capacitance arms were used to block the d.c. paths and the measurements were independent of the frequency. The impedance of the arc path being low, the effect of the stray capacitances was negligible. The main difficulty was due



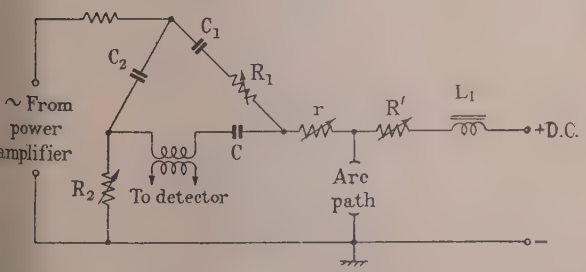
ig. 9.—Change of electron concentration along the radius of the tube container.

Arc current .. .. .. 1·5 amp
Bath temperature .. .. .. 40° C



g. 10.—Difference in potential between points along the radius and the axis of the tube container.

Arc current .. .. .. 1·5 amp
Bath temperature .. .. .. 40° C



ig. 11.—Circuit used for the measurement of the effective impedance of the arc path.

R′ = Limiting resistor to change the value of the direct arc current.
L_1 = Choke to eliminate the shunting effect of the d.c. supply.
r = Small variable resistor in series with the arc path in the a.c. circuit.
C = Blocking condenser.

ne noise voltage appearing between the arc electrodes. The etector was designed to be an amplifier with a bridge-output ircuit, one arm of which was a parallel-tuned circuit tuned to ne operating frequency. The out-of-balance noise voltage is ed back into the input, thus increasing the detector selectivity.

The results of measurements in each group of tests are given in Tables 1 and 2.

Table 1

RESULTS OF MEASUREMENTS FOR 20° C BATH TEMPERATURE

Arc current	Electric field	Electron temperature	R_{150}/R_{75}	L_{150}/L_{75}
amp	volt/cm	°K		
4	0·263	21 750	1·78	2·1
2·5	0·295	21 800	1·58	2·1
1·5	0·31	23 800	1·8	1·8

Table 2

RESULTS OF MEASUREMENTS FOR 40° C BATH TEMPERATURE

Arc current	Electric field	Electron temperature	R_{150}/R_{75}	L_{150}/L_{75}
amp	volt/cm	°K		
4	0·31	15 200	1·86	1·88
2·5	0·365	14 900	2	1·54
1·5	0·37	14 300	1·7	1·85

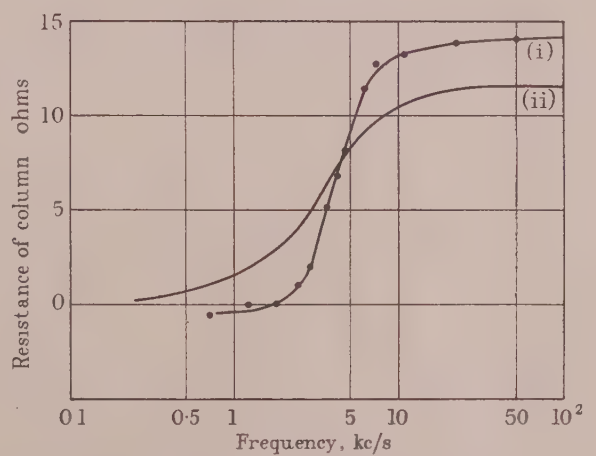


Fig. 12A.—Effective resistance of the 150cm positive column.

Arc current .. .. .. 4 amp
Bath temperature .. .. .. 40° C
(i) Experimental.
(ii) Calculated.

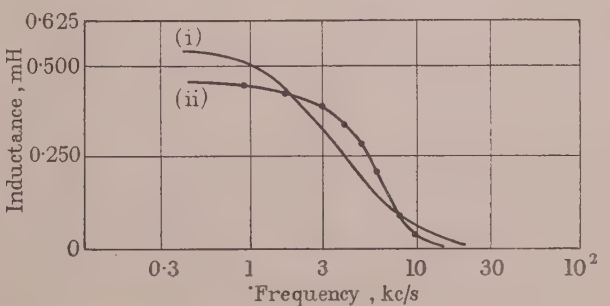


Fig. 12B.—Effective inductance of the 150cm positive column.

Arc current .. .. .. 4 amp
Bath temperature .. .. .. 40° C
(i) Calculated.
(ii) Experimental.

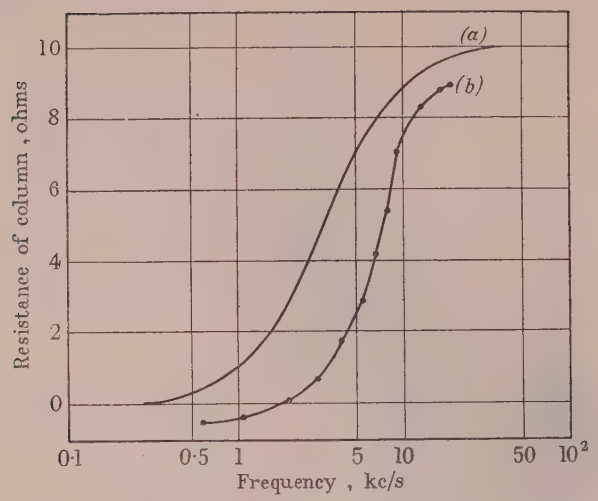


Fig. 13.—Effective resistance of the 150 cm positive column.

Arc current .. .. .. 4 amp
Bath temperature .. .. 20° C
(*a*) Calculated.
(*b*) Experimental.

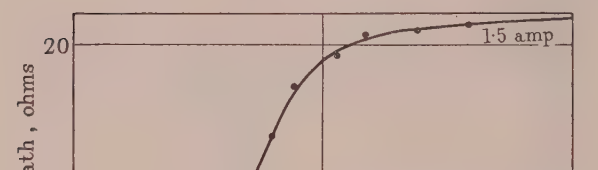


Fig. 14.—Effective resistance of different arc currents in the 75 cm discharge tube.

Bath temperature .. .. 40° C

The values given for the electric field and electron temperatures are the average of the readings obtained in the two tubes. Actually these readings were quite close to each other, the maximum difference being not more than 3%. Again, the average ratios, at any frequency, of the resistance and inductance of the 150 cm and 75 cm columns are given as R_{150}/R_{75} and L_{150}/L_{75} respectively. In Figs. 12A and 12B the resistance and inductance/frequency curves (obtained experimentally), for a 4-amp direct arc current in a 150 cm tube at a bath temperature of 40° C, are compared with the calculated curves under the same conditions. Figs. 14 and 15 give the measured resistance and inductance/frequency curves for the three different currents.

(4.3) Discussion and Conclusions

(4.3.1) General.

The change of the arc-path impedance with frequency is due to the changes of electron concentration being dependent on the frequency. At low frequencies, the periodic time is long enough

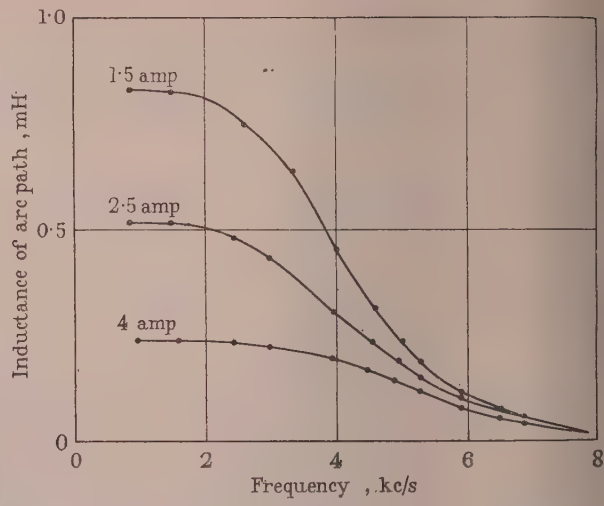


Fig. 15.—Effective inductance of different arc currents in the 75 cm discharge tube.

Bath temperature .. .. 40° C

for all the particles capable of doing so to take part in the ioniz tion process, and thus a large change in electron concentratio and a low impedance of the path are produced. At high fr quencies, however, the changes of the electric field are so rapi that only a small number of ionizing particles will take part the ionization process. In this case, practically no change in th electron concentration is produced during the cycle, and hen a constant and relatively high resistance of the path results.

Both experimental and calculated curves of the high- an low-pressure cases agree well in this respect. The resistan curves start from a low value at low frequencies and change a constant and relatively large value at high frequencies. Furthe more, the frequencies at which the bends in the curves occur a very close in both the experimental and calculated results.

(4.3.2) High-pressure Column.

(*a*) The experimental work done by Magnold is not sufficie to check the theory developed. Further work is necessary investigate the effect of changing the length, pressure and a current in the column in order to make an exact compariso However, the calculated curves are seen to have the same sha and order of magnitudes as the experimental observations Magnold.

(*b*) The above theory applies only for the positive column the arc, while the resistances obtained by Magnold were measur between the electrodes, i.e. including the cathode region. Buchtiger's and the authors' investigations of the behaviour the hot-cathode region of an arc of very short length, it was fou that in the range of frequency considered above, the arc had nearly constant positive resistance while its effective inductan was negligible. Therefore the measured resistances of the a path will be the sum of the resistances of the positive column a the cathode in series. This may explain why the measur resistance starts from a positive value at low frequencies and r from a negative one, as indicated by the calculated curve.

(*c*) In order to make the calculated resistance/frequency cur fall in the same frequency range as the experimental one it w necessary to take the probability of thermal ionization F equal unity, which is relatively high.

(*d*) The theory developed above is mainly based on the Sa formula for the concentration of electrons and the assumpti that the ionization is a thermal process. The theory will r

ld if the pressure is reduced to a value where the electron ization predominates. However, for higher pressures when e Saha formula is valid, the theory holds.

(*e*) The calculations are carried out for an arc in mercury pour which is a mono-atomic gas, and no dissociation takes ace. For other gases, the dissociation phenomena should be nsidered. Further, in the case of mercury vapour the radiation ectrum can be considered uniform, which may not be true for her gases.

(*f*) The method of treatment given above bears a strong nilarity to that given by Kneser[15] to explain the behaviour of und absorption in gases. The ionization process and the change of the ionization energy between neutral molecules in e positive column are analogous to the exchange of the vibra- onal energy between the gas particles in the case of sound.

3.3) **Low-Pressure Column.**

(*a*) The frequency at which the resistance/frequency curve arts to bend will depend upon the value of $A = 3/2T(d\alpha/dT)$. he resistance/frequency curve will be shifted towards the higher- equency region, the smaller the value of A. For a bath tem- erature of 40° C, $A = 1{\cdot}06 \times 10^{-4}$ and for a bath temperature 20° C, $A = 0{\cdot}8 \times 10^{-4}$. The resistance curve for a bath mperature of 20° C should be to the right of that at 40° C for e same arc current.

The theoretical deductions agree to a certain extent with the perimental findings. The shift in the experimental curve is ore than expected, as shown in Fig. 13.

(*b*) The effective resistance at high frequency is given by $\mathscr{E}/\Delta I = \mathscr{E}/I =$ constant. As the electric field in the column aries very little with the direct arc current for the same bath mperature the resistance of the path should be less for higher rrents. The agreement in this respect is seen from Figs. 14 d 15. The same argument will hold for the expression for the ductance.

(*c*) The theoretical expression for the effective impedance was erived per unit length of the column. Therefore the impedance the 150 cm tube should be double that of the 75 cm one. It ll be seen from Tables 1 and 2 that the ratio of the two im- dances ranges from 1·5 to 2·1. The agreement is fairly satis- ctory. The departure from a 2 : 1 ratio may be due to the ct that the two tubes have a common path between the surface the cathode pool and the neck where they branch out. Further- ore, the effect of the phenomena in the cathode region was not ken into account. In order to minimize this effect, the positive lumn under consideration was made of considerable length. owever, about 30% of the total voltage between the electrodes isted across the cathode region.

(5) REFERENCES

(1) Reich, H. J.: "Theory and Application of Electron Tubes" (McGraw-Hill, 1944).

(2) Von Weizel, Rompe, and Schulz: "Zur theorie der Modulation eines Hochdruckbogens," *Zeitschrift für Technische Physik*, 1940, **12**, p. 387.

(3) Cobine, J. D.: "Gaseous Conductors" (McGraw-Hill, 1941).

(4) Killian, T. J.: "The Uniform Positive Column of an Electric Discharge in Mercury Vapour," *Physical Review*, 1930, **35**, p. 1238.

(5) Langmuir, I., and Mott-Smith, H.: "Studies of Electric Discharges in Gases at Low Pressures," *General Electric Review*, 1924, **27**, pp. 449, 538, 762 and 876.

(6) Maxfield, F. A., and Benedict, R. R.: "Theory of Gaseous Conduction and Electronics" (McGraw-Hill, 1941).

(7) Buchtiger, P.: "Die Dynamischen Charaktaristiken einer Bogentladung," *Helvetia Physica Acta*, 1920, **3**, p. 335.

(8) Magnold, H.: *Elektrotechnische Nachrichten*, 1940, **17**, p. 57.

(9) Druvestyn and Penning: "Electrical Discharges in Gases," *Review of Modern Physics*, 1939, **12**, p. 87.

(10) Ornstein, L. S., and Brinkman, H.: "Der Thermische Mechanismus in der Säule des Lichtbogens," *Physica*, 1934, **1**, p. 797.

(11) Ellenbass, W.: "Der Gradient der Überhochdruck-Quecksilber Entladung," *ibid.*, 1937, **4**, p. 279.

(12) Suits, C. G.: *Physics*, 1930, **6**, pp. 190 and 315.

(13) Suits, C. G.: "Current Densities, Lumen Efficiency and Brightness in A, N_2, He and H_2 Arcs," *Journal of Applied Physics*, 1939, **10**, p. 730.

(14) Von Engel, A., and Stienbeck, M.: "Elecktriche Gastentladungen."

(15) Kneser, H. O.: "The Interpretation of the Anomalous Sound-Absorption in Air and Oxygen in terms of Molecular Collisions," *Journal of the Acoustical Society of America*, 1933–34, **5**, p. 122.

(16) Compton, K. T.: *Review of Modern Physics*, 1930, **12**, p. 211.

(6) APPENDICES

(6.1) Effect of Diffusion alone on the Electron Concentration

Considering a cylindrical core of radius a with a uniform concentration of electrons N_1 per cubic centimetre, the concentration at any point at radius r outside the column will be

$$\frac{d^2N}{dr^2} + \frac{1}{r}\frac{dN}{dr} = \frac{1}{D}\frac{dN}{dt}$$

For alternating concentrations

$$\frac{d^2N}{dr^2} + \frac{1}{r}\frac{dN}{dr} - \frac{j\omega}{D}N = 0$$

and $$n = N = A\mathrm{K}_0 r\sqrt{(j\omega/D)} + B\mathrm{I}_0 r\sqrt{(j\omega/D)}$$

where A and B are constants and K_0 and I_0 are Bessel functions of modified forms. As the concentration is finite at a very large radius, $B = 0$.

Again, $N = N_1$ at $r = a$, and therefore $A = N_1/\mathrm{K}_0 a\sqrt{(j\omega/D)}$. The rate of diffusion from the core per centimetre length will be

$$\Delta Q = -2\pi a D\left(\frac{dn}{dr}\right)_{r=a}$$

Using the asymptotic form

$$\mathrm{K}_0(x) = \sqrt{\left(\frac{\pi}{2x}\right)}\epsilon^{-x}$$

The amplitude Δq of the number of electrons which alternates outwards and inwards at the boundary surface $r = a$ will be

$$\Delta q = \Delta Q/j\omega = \frac{\pi D N_1}{j\omega}[1 + 2a\sqrt{(j\omega/D)}]$$

To maintain the concentration n_1 in the core, electrons have to be introduced at a rate equal to the sum of the rate of increase of the electrons in the core, and the rate of their loss at the boundary $r = a$. Thus

$$j\omega(\Delta\xi) = j\omega\pi a^2 N_1 + \Delta Q$$

$$\frac{\Delta q}{\Delta\xi} = D[1 + 2a\sqrt{(j\omega/D)}]\Big/j\omega\left\{a^2 + \frac{D}{j\omega}[1 + 2a\sqrt{(j\omega/D)}]\right\}$$

(6.2) Effect of Ionization and Diffusion combined on the Electron Concentration in the Low-Pressure Column

The electron concentration is defined by Schottky to be

$$\frac{d^2N}{dr^2} + \frac{1}{r}\frac{dN}{dr} + \frac{\alpha}{D}N = \frac{1}{D}\frac{dN}{dt}$$

For the d.c. column $dN/dt = 0$ and $N = N_0 J_0 r\sqrt{(\alpha/D)}$.

If the electron temperature is increased by a small sinusoidal quantity $\Delta T\epsilon^{j\omega t}$, the coefficient of ionization, α, and the electron concentration, N, will undergo similar sinusoidal changes, $\Delta\alpha\epsilon^{j\omega t}$ and $n\epsilon^{j\omega t}$, respectively. Assuming that the coefficient of diffusion D remains constant, the new concentration will be

$$\frac{d^2}{dr^2}(N + n\epsilon^{j\omega t}) + \frac{1}{r}\frac{d}{dr}(N + n\epsilon^{j\omega t}) + \frac{(N + n\epsilon^{j\omega t})(\alpha + \Delta\alpha\epsilon^{j\omega t})}{D} = \frac{1}{D}\frac{d}{dt}(N + n\epsilon^{j\omega t})$$

Subtracting the steady-state relation $dN/dt = 0$, and neglecting the term of second order,

$$\frac{d^2n}{dr^2} + \frac{1}{r}\frac{dn}{dr} + \frac{\alpha - j\omega}{D}n + \frac{\Delta\alpha}{j\omega}N_0 J_0 r\sqrt{(\alpha/D)} = 0$$

The complementary function

$$\frac{d^2n}{dr^2} + \frac{1}{r}\frac{dn}{dr} + \frac{\alpha - j\omega}{D}n = 0$$

will have the solution

$$n = CJ_0 r\sqrt{[(\alpha - j\omega)/D]}$$

Taking $J_0 r\sqrt{(\alpha/D)}$ as a particular integral, the complete solut will be

$$n = CJ_0 r\sqrt{[(\alpha - j\omega)/D]} + \frac{\Delta\alpha}{j\omega}N_0 J_0 r\sqrt{(\alpha/D)}$$

The constant C can be determined from the boundary condit $n = 0$ at $r = a$ as the wall is an insulator; therefore $C =$ The change in the number of electrons per centimetre length the column is

$$\Delta\eta = \int_0^a 2\pi r n dr = \frac{\Delta\alpha}{j\omega}\int_0^a 2\pi r N_0 J_0 r\sqrt{(\alpha/D)}dr$$

$$= \frac{\Delta\alpha}{j\omega}\eta \qquad = \frac{\eta(d\alpha/dT)}{j\omega}\Delta T$$

.318.42: 621.3.013

Monograph No. 102 M
July 1954

THE DESIGN OF COILS FOR THE PRODUCTION OF HIGH MAGNETIC FIELDS

By A. N. INCE, B.Sc.

(*The paper was first received* 3*rd March, and in revised form* 28*th April*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in July*, 1954.)

SUMMARY

The paper is concerned with the design of coils of rectangular cross-tion for the production of intense transient magnetic fields using rgy obtained from a bank of charged condensers. Curves are given facilitate the design of a coil of maximum efficiency to meet any en requirements.

LIST OF PRINCIPAL SYMBOLS

r_2 = Internal and external radii of the coil, cm.
$2l$ = Total length of the coil, cm.
N = Total number of turns on the coil.
ρ = Resistivity of coil wire, ohm-cm.
R = Resistance of the coil, ohm.
L = Inductance of the coil, henry.
λ = Space factor.
v = Conductor volume, cm^3.
κ = Nagaoka constant.
c = Specific heat of the coil wire, cal/g/°C.
d = Density of the coil wire, g/cm^3.
i = Instantaneous current, amp.
$_{max}$ = Peak current, amp.
$_{max}$ = Maximum field intensity at the centre of the coil, oersteds.
τ = Duration of the first half-cycle of current, sec.
C = Capacitance of condenser, farads.
V = Voltage to which the condenser is charged, volts.
W = Energy stored in the condenser, joules.

(1) INTRODUCTION

The production of very high transient magnetic fields has en described by Kapitza,[1] who made use of energy stored in generator, and the design of coils for use with this method has en treated by Cockcroft.[2] More recently a similar method has en developed in which the energy is obtained from a bank of arged condensers,[3] and Myers[4] has given some constructional ails of suitable coils, about 2cm long and 0·64 to 1·73cm in ernal diameter, capable of producing magnetic fields up to $\times 10^3$ oersteds with periods of the order of 50 microsec. general design procedure is given.

The author has recently employed this method in connection h some work on electro-luminescent powders, and was tunate in being able to make use of apparatus which had viously been installed in the Royal Society Mond Laboratory quite different experiments. It was necessary, however, to ign some new coils, and the present paper is concerned with nsiderations which arose in this connection. Curves are given facilitate the design of a coil of maximum efficiency to meet en requirements with regard to period, intensity of field, and gth and internal radius of coil. The method of design is an ension of that previously used by Shoenberg.[5]

(2) DESIGN PROCEDURE

(2.1) Preliminary Considerations

The following discussion is limited to the case where the discharge from the condenser is oscillatory, as it will be in nearly all practical cases. Very often the Q-factor of the circuit will be of the order of unity, and when this is the case, the connection between condenser and coil may be established by means of any convenient high-speed switch, for example a gravity-operated switch in which a heavily weighted metallic point falls at high velocity into a block of lead. Alternatively, if the Q-factor is not very low, the discharge may be accomplished by means of an ignitron, so that only the first half-cycle of current is permitted to flow. In either case the heating of the coil can be assumed to result only from energy dissipated during the first half-cycle, and this assumption will be made throughout.

The factors affecting the design of a coil for any particular purpose then arise in the following order:

The internal radius of the coil will normally be determined by conditions over which the designer has little control, although it is obvious that this radius should be as small as possible.

Similarly, external conditions will probably determine the minimum value that can be assigned to the duration of the first half-cycle of the discharge; although here again, the smaller the value permitted the greater will be the magnetic field that can be obtained, other factors remaining the same.

Once the internal radius and the period of oscillation have been settled, curves given later enable us to find the best shape of coil, the best gauge of wire and the optimum number of turns for a given capacitance. It then becomes possible to calculate the maximum field to be obtained from a given stored energy, or the minimum stored energy for a given field.

Finally, the configuration of the field can be estimated. If it is not sufficiently uniform, a longer coil containing the same volume of copper must be used, with a corresponding sacrifice in the maximum value of the magnetic field.

The above discussion assumes that the heating of the coil is the limiting factor, but if the coil has a relatively high Q-factor, mechanical stresses may become important. No attempt is made to calculate these stresses because it is not easy to set an upper limit to the stress which a coil of any particular construction will withstand. Thus the most straightforward procedure is to construct the optimum coil, taking temperature rise to be the limiting factor, and to determine by experiment whether this coil is strong enough to withstand the mechanical stresses. If it is not, the stresses can be reduced by using a longer coil with the same volume of copper, but such a coil will produce a smaller maximum field.

(2.2) Explanation of the Design Curves

In this Section the detailed procedure for designing a coil is described, making use of a number of curves. The derivation of the formulae on which these curves are based is straightforward and is given in Section 5.

orrespondence on Monographs is invited for consideration with a view to lication.
r. Ince is in the Department of Engineering, University of Cambridge.

Consider a coil of length $2l$, with internal and external radii r_1 and r_2, respectively. It is convenient to express the size and shape of this coil in terms of the two parameters

$$a = \frac{r_2}{r_1} \text{ and } b = \frac{l}{r_1}$$

r_1 is assumed to be fixed by the purpose for which the coil is to be used, and so also is the duration τ of the first half-cycle of the discharge.

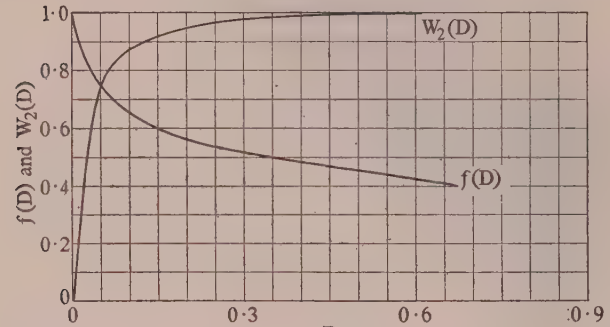


Fig. 1.—Variations of the functions f and W_2 with D.

From the known values of r_1 and τ, a parameter D given b

$$D = 1/[1 + (r_1^2/100\tau)^2] \quad . \quad . \quad . \quad .$$

is calculated. D is, in fact, inversely proportional to the squ of the Q-factor of the coil.

The maximum field strength H_{max} at the centre of the can then be expressed as

$$H_{max} = 2{\cdot}52 \times 10^4 \, f(D)\eta(a, b)\sqrt{(W/r_1^3)} \text{ oersteds} \quad .$$

where W is the energy in joules stored in the condenser, f(D a function of D which is plotted in Fig. 1, and $\eta(a, b)$ is a fu tion of a and b which determines the efficiency of the c Curves of constant η are plotted in Fig. 2, and the determinat of η will be considered below.

The energy dissipated in the first half-cycle is given appr mately by

$$W_1 = W\{1 - \exp[-2\pi\sqrt{[D/(1 - D)]}\} \quad . \quad .$$
$$= W W_2(D)$$

where $W_2(D)$ is plotted as a function of D in Fig. 1. Since can be calculated from the capacitance and voltage rating the condenser which is available, W_1 can be determined.

The quantity of heat lost from the wire of the coil dur time τ will be quite negligible, so the volume v of the conducto related to its density d, specific heat c and the maximum p missible temperature rise ΔT by

$$v = W_1/4{\cdot}18dc\Delta T \quad . \quad . \quad . \quad . \quad .$$

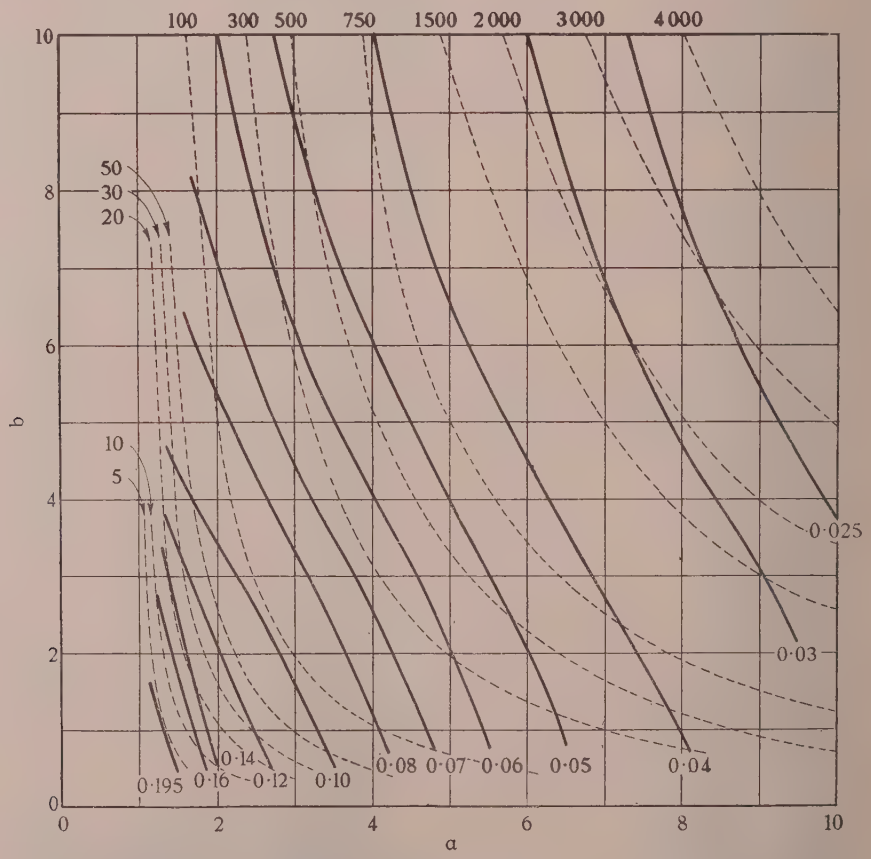


Fig. 2.—Efficiency, η, and volume, v_1, curves as a function of the coil parameters a and b.

——— η.
----- v_1.

λ is the space factor, i.e. the fraction of the total volume of the il occupied by metallic conductor,

$$v = \lambda r_1^3 \pi b(a^2 - 1) = \lambda r_1^3 v_1(a, b) \quad . \quad . \quad . \quad . \quad (5)$$

here the function v_1 is proportional to v, and curves of constant are plotted in Fig. 2.

In eqn. (4) the product dc is approximately equal to 0·84 for y of the materials likely to be used for the conductor (e.g. pper or cadmium–copper), and so substitution from eqn. (5) ves

$$v_1 = W\mathrm{W}_2/3{\cdot}5\lambda r_1^3 \Delta T \quad . \quad . \quad . \quad . \quad . \quad (6)$$

is unlikely to be very different from 0·6, and this value is sumed in the first instance. Thus, putting ΔT equal to some lue between 100 and 150° C, depending on the construction the coil and the type of insulation, v_1 can be calculated from n. (6).

From the constant-v_1 and constant-η curves plotted in Fig. 2, is possible to find the maximum value of η that can be obtained th the value of v_1 fixed as above. The same curves also show at values of a and b must be used to attain this value of η. bstitution of η in eqn. (2) gives the maximum field H_{max}. If $_{max}$ is not as large as is required, W must be increased by the e of a larger condenser and the calculations repeated.

Once a and b have been settled, the number of turns can be termined from

$$^2 = \frac{8r_1^3\lambda^2 D}{10^9 C\rho^2} \cdot \frac{b(a-1)^2}{1 + 0{\cdot}225(1+a)/b + (a-1)[0{\cdot}64/(1+a) + 0{\cdot}42/b]} \quad . \quad (7)$$

here the capacitance, C, of the condenser is in farads and is the mean resistivity of the wire over the appropriate mperature range. For example, ρ may be assumed to be 2×10^{-6} ohm-cm for copper, with $\Delta T = 130°$ C and the coil tially at room temperature.

From N and the dimensions of the coil, the appropriate gauge wire may be determined, and a check can then be made of ether the assumed value of space factor, $\lambda = 0{\cdot}6$, is correct. not, the calculations must be repeated with a revised value of λ. Since the efficiency of the coil increases with λ, every effort ust be made to keep λ as large as possible. For wire larger an, say, No. 10 S.W.G., the use of square-section conductor ould be worth while.

(2.3) Results obtained from the above Procedure

The following results are quoted as examples of the above sign procedure:

In Fig. 3, the variation of H_{max} with r_1 is shown for $W =$ 00 joules, $D = 0{\cdot}05$ and $\Delta T = 100°$ C. Similarly, Fig. 4 ows the variation of H_{max} and τ with D, for $W = 5\,000$ joules, $T = 100°$ C and $r_1 = 0{\cdot}9$cm.

If there were no limitation due to overheating of the coil, e value of H_{max} obtainable with a given coil would be pro- rtional to the square root of W, other factors remaining the me. As it is, any increase in W involves an increase in the lume of the coil to avoid overheating, and so H_{max} varies atively slowly with W. In a particular case ($r_1 = 0{\cdot}9$cm and $= 0{\cdot}05$) an increase of W from 5 000 to 10 000 joules causes ly a 16% increase in H_{max}.

An experimental coil was constructed to give a maximum field $1{\cdot}3 \times 10^5$ oersteds with $r_1 = 0{\cdot}9$cm, $W = 5\,000$ joules, $T = 100°$ C and $D = 0{\cdot}4$. The maximum current through coil was measured by observing with an oscillograph the tage drop across a small resistance in series with the coil. is current agreed to within 3% with the value obtained from formulae previously given.

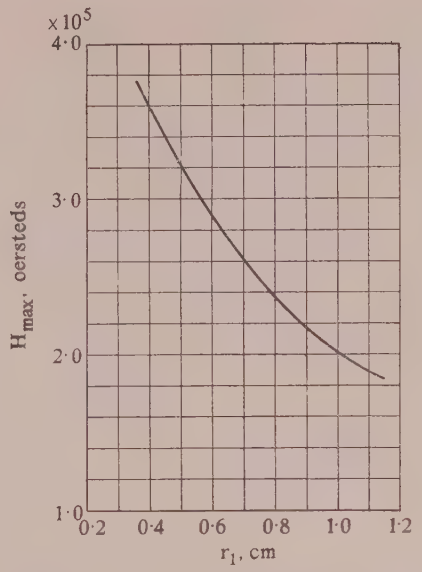


Fig. 3.—Dependence of maximum field intensity H_{max} on the internal radius, r_1, of the coil for $W = 5\,000$ joules, $D = 0{\cdot}05$, $\Delta T = 100°$ C.

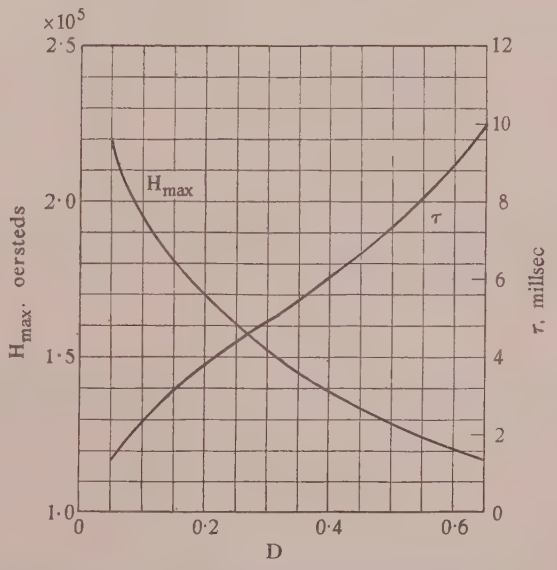


Fig. 4.—Dependence of maximum field intensity H_{max} and period τ of field on D, for $W = 5\,000$ joules, $r_1 = 0{\cdot}9$cm, $\Delta T = 100°$ C.

(3) ACKNOWLEDGMENTS

The author wishes to express his thanks to Dr. D. Shoenberg, who kindly allowed him to carry out experimental work in the Royal Society Mond Laboratory and to make use of apparatus previously constructed there for producing high magnetic fields. He is also indebted to Mr. E. Laurmann, who assisted him with the experimental work.

(4) REFERENCES

(1) KAPITZA, P.: *Proceedings of the Royal Society*, A, 1927, **115**, p. 658.

(2) COCKCROFT, J. D.: "The Design of Coils for the Production of Strong Magnetic Fields," *Philosophical Transactions of The Royal Society of London*, A, 1928, **227**, p. 317.
(3) SHOENBERG, D.: *Nature*, 1952, **170**, p. 569; and *Physica*, 1953, **19**, p. 791.
(4) MYERS, W. R.: "The Construction of Small Solenoids for the Production of Intense Magnetic Fields," *Journal of Scientific Instruments*, 1953, **30**, p. 237.
(5) SHOENBERG, D.: Private communication.
(6) WELSBY, V. G.: "The Theory and Design of Inductance Coils" (Macdonald and Co., Ltd., London, 1950), p. 25.

(5) APPENDIX

For a coil with the dimensions given in Section 2.2, the field at the centre is readily shown to be

$$H = \frac{\pi Ni}{5(r_2 - r_1)} \log_e \frac{r_2 + \sqrt{(r_2^2 + l^2)}}{r_1 + \sqrt{(r_1^2 + l^2)}} \text{ oersteds} \quad . \quad (8)$$

where i is the current in amperes.

The resistance R of the coil is

$$R = \frac{\pi\rho N^2(r_1 + r_2)}{2l\lambda(r_2 - r_1)} \quad . \quad . \quad . \quad . \quad . \quad (9)$$

so that eqn. (8) may be written

$$H = \frac{i}{5} \frac{\sqrt{(2\pi l\lambda R)}}{\sqrt{[\rho(r_2^2 - r_1^2)]}} \log_e \frac{r_2 + \sqrt{(r_2^2 + l^2)}}{r_1 + \sqrt{(r_1^2 + l^2)}} \quad . \quad (10)$$

or in terms of a and b,

$$H = \frac{i}{5} \sqrt{\frac{\lambda R}{r_1\rho}} \sqrt{\frac{2\pi b}{a^2 - 1}} \log_e \frac{a + \sqrt{(a^2 + b^2)}}{1 + \sqrt{(1 + b^2)}}$$

$$= i\sqrt{\frac{\lambda R}{r_1\rho}} \mathrm{G}(a, b), \text{ say} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Where G is a function of a and b previously introduced by Cockcroft.[2]

If a condenser of capacitance C, charged to a voltage V, is connected to the coil of inductance L at time $t = 0$, the subsequent current in the coil is given by

$$i = \frac{V}{L\omega} \varepsilon^{-\mu t} \sin \omega t \quad . \quad . \quad . \quad . \quad . \quad (12)$$

where $\mu = R/2L$, $\omega^2 = (1/LC) - (R^2/4L^2)$ and the equation is valid for $\omega^2 > 0$. The maximum current during the first half-cycle is then

$$I_m = \frac{V}{L\sqrt{(\mu^2 + \omega^2)}} \exp\left(-\frac{\mu}{\omega} \arctan \frac{\omega}{\mu}\right) . \quad . \quad (13)$$

It is convenient to rewrite eqn. (13) in terms of a new variable $D = R^2C/4L$, to give

$$I_m = \sqrt{(C/L)}V \exp\left\{-\sqrt{[D/(1 - D)]} \arctan \sqrt{[(1 - D)/D]}\right\} \quad . \quad . \quad . \quad . \quad (14)$$

$$= V\sqrt{(C/L)}\mathrm{f}(D) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

The curve of f(D) against D in Fig. 1 is computed from th equations.

The inductance L of a coil of the form considered is given[6]

$$L = 10^{-9}\pi^2 r_1 N^2 \kappa(1 + a)^2/2b \text{ henrys} \quad . \quad . \quad (16)$$

where

$$\kappa = 1\Big/\left[1 + \frac{0{\cdot}225(1 + a)}{b} + (a - 1)\left(\frac{0{\cdot}64}{1 + a} + \frac{0{\cdot}42}{b}\right)\right] \quad (17)$$

Eqns. (9) and (16) can be used to express D in the form

$$D = 10^9 C\rho^2 N^2/8\kappa b r_1^3 \lambda^2 (a - 1)^2$$

which is identical with eqn. (7).

Since the energy stored in the condenser is $\frac{1}{2}CV^2$, eqns. ((15) and (16) can be combined to give

$$H_{max} = 2{\cdot}52 \times 10^4 \sqrt{(W/r_1^3)}\mathrm{f}(D)\mathrm{G}(a, b)/\sqrt{[\kappa(a^2 - 1)]}$$

which is equivalent to eqn. (2) if $\eta(a, b)$ is written instead $\mathrm{G}(a, b)/\sqrt{\kappa}\sqrt{(a^2 - 1)}$. The curves of $\eta(a, b)$ plotted in Fig were computed from this relation.

The half-period τ of the first oscillation is found by putt $\omega\tau = \pi$ in eqn. (12). Whence

$$\tau = \pi\Big/\sqrt{\left(\frac{1}{LC} - \frac{R^2}{4L^2}\right)}$$

$$= \frac{2\pi^2 \lambda r_1^2 \kappa(a^2 - 1)}{10^9 \rho} \sqrt{\frac{D}{1 - D}} \quad . \quad . \quad . \quad (18)$$

For purposes of design an equation is required from which can be calculated when r_1 and τ are given. Eqn. (18) as it sta cannot be used for this purpose, because the values of λ, κ an will not be known at this stage of the design. However, λ not differ greatly from 0·6, and it will usually be sufficien accurate to put $\kappa(a^2 - 1) = 2$. If, in addition, ρ is given value $2{\cdot}2 \times 10^{-6}$ ohm-cm, which is appropriate to copper w $\Delta T = 130°$ C, eqn. (18) reduces to the approximate form

$$D = 1/[1 + (r_1^2/100\tau)^2]$$

which is identical with eqn. (1).

For the calculation of the energy dissipated during the f half-cycle, a mean value of the resistance R over the appropri temperature range may be used. Then

$$W_1 = R\int_0^{\tau} i^2 dt = \frac{RV^2}{L^2\omega^2}\int_0^{\pi/\omega} \epsilon^{-2\mu t} \sin^2 \omega t dt$$

which can be reduced to

$$W_1 = W\left\{1 - \exp\left[-2\pi\sqrt{(D/(1 - D)}\right]\right\}$$

$$= W\mathrm{W}_2(D)$$

$\mathrm{W}_2(D)$ is plotted against D in Fig. 1.

.315.5.066.6 : 621.3.041.1

Monograph No. 103
July 1954

THE VARIATION WITH CURRENT AND INDUCTANCE OF METAL TRANSFER BETWEEN PLATINUM CONTACTS

By JANET RIDDLESTONE, B.A.

The paper was first received 5th March, and in revised form 30th April, 1954. It was published as an INSTITUTION MONOGRAPH *in July, 1954.)*

SUMMARY

The investigation described is a continuation of the work on transfer ween platinum contacts published by the author in a previous paper. account is given of the manner in which the metal transfer between tinum contacts breaking a 6-volt circuit varies with current and cuit inductance. Curves of the net transfer are given for currents d inductances in the ranges 1·8–7·6 amp and 0·06–117 μH pectively. The net transfer consists of up to four different types of nsfer superimposed one on another, namely "bridge," "short arc," ng arc" and "reversed short arc" transfers. The existence of the rth type of transfer had been observed previously by Dr. A. L. Allen currents of 10 and 20 amp, but was not known to occur at currents low as 2 amp. It is shown that the build-up of pips on the contacts ies in steepness with the nature of the transfer. Some suggestions given as to the mechanism of the different types of transfer, but at sent a complete explanation cannot be given. Consideration is en to the practical application of the results, and it is shown that der some conditions the life of platinum contacts could be improved controlling the effective circuit inductance at break to a value about 0·6 μH.

(1) INTRODUCTION

In a previous paper by the author[1] it was shown that "fine nsfer" between platinum contacts breaking low-current low-ltage circuits occurs as residual transfer, which is independent circuit inductance, together with a "short-arc transfer" which reases with circuit inductance. Both types of transfer result a pip build-up on the cathode and a corresponding crater on anode. These results were obtained with a 6-volt supply and sed-circuit currents of 0·95 and 1·75 amp and controlled luctances of 0·05–10 μH. In order to obtain further informa-n about transfer, the ranges of currents and inductances have en extended, and results are given in the paper for currents 1·8, 3·0, 3·9 and 7·6 amp and inductances from 0·06 to 7 μH. The results confirm those given in the earlier paper, t they also show that the variation of transfer with inductance nore complex at higher currents.

(2) TEST APPARATUS AND CIRCUIT

Except for a change in the contact-operating mechanism, the perimental arrangement was essentially the same as that cribed in the previous paper. The distributor unit used to erate the contacts was replaced by a magnetically operated y designed to give a controlled contact pressure and to have opening and closing speeds. The moving contact was unted on an armature which was pivoted so as to move in the between two coil-operated magnets. One coil provided the hold-on force and the other provided the a.c. operating force. armature was suspended by means of steel shafts in bearings igned to eliminate play at the contacts. The amount of play ich can be tolerated in any direction at right angles to the motion of the armature is determined by the size of the area over which transfer occurs. For platinum contacts breaking 1·8 amp the maximum play allowable is probably about 5×10^{-4} cm. The relay operated satisfactorily for currents of 1·8 amp upwards, but would not give accurate results at lower currents owing to small movements at the contact area. The contacts were crossed cylinders of 22 s.w.g. thermo-pure platinum. The contact pressure was 30 g, the average opening speed was about 0·5 cm/sec, and the frequency of operation was 50 c/s. An auxiliary relay was arranged so that the platinum contacts broke the circuit current but closed on zero current.

The circuit was designed so that the inductive energy available at the contacts could be controlled down to the smallest values possible. A diagram of the circuit is shown in Fig. 1(*a*). The

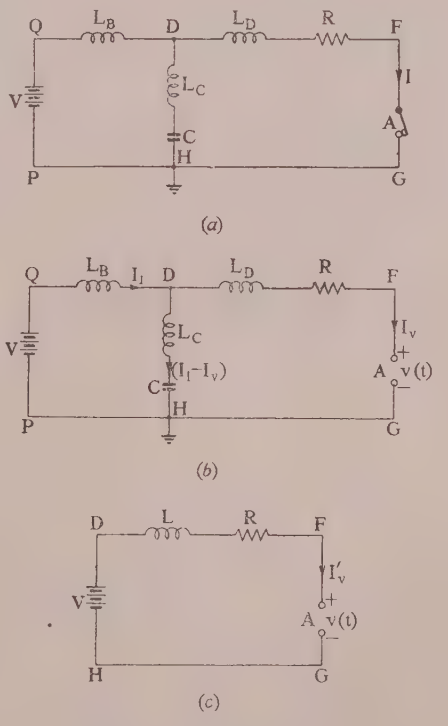


Fig. 1.—Circuit diagrams.
(*a*) With contacts closed.
(*b*) After an arc is established.
(*c*) Equivalent circuit for the arc current.

capacitor C acts as a by-pass for the inductive energy in the battery leads. The self-inductance, L, of the loop DFGH is defined as the sum of the self-inductance, L_0, of the loop together with a controlled amount of added inductance. The value of L_0 varies by a small amount with the current-limiting resistance R. This resistance is in the form of a suitable combination of 6-volt

rrespondence on Monographs is invited for consideration with a view to ication.

rs. Riddlestone is with the British Electrical and Allied Industries Research ciation.

car lamps which have low self-inductance and small dimensions. The values of the parameters are as follows:

Battery voltage, $V = 6$ volts.
Self-inductance of battery leads, $L_B \simeq 0{\cdot}33\ \mu$H.
Self-inductance of capacitor + leads, $L_C = 0{\cdot}01\ \mu$H.
Inductance of loop DFGH, $L = L_C + L_D = 0{\cdot}06$ to $117\ \mu$H.
Residual inductance of DFGH, $L_0 = 0{\cdot}06$ to $0{\cdot}1\ \mu$H.
Capacitance, $C = 100\ \mu$F.
Current-limiting resistance, $R = 0{\cdot}79, 1{\cdot}54, 2$, or $3{\cdot}33\ \Omega$.
Closed-circuit current, $I = 7{\cdot}6, 3{\cdot}9, 3{\cdot}0$, or $1{\cdot}8$ A.
Current just before break, $I_B = 5{\cdot}6, 2{\cdot}9, 2{\cdot}2$, or $1{\cdot}3$ A.

The inductances were measured with the circuit in the position occupied during the tests at frequencies between 2 and 26 Mc/s.

It is well known[1–4] that as the contacts open the voltage across them rises to a value V_B (=1·6 volt approximately) before complete separation occurs. Calculations show[1] that if no arcing takes place the voltage across the contacts immediately after break is given by

$$v(t) = I_B\sqrt{\left(\frac{L}{C_0}\right)}\sin\omega t + V_B \quad . \quad . \quad . \quad (1)$$

where C_0 ($\simeq 10\ \mu\mu$F) is the capacitance at the contacts immediately after break and ω^2 is equal to $1/C_0L$. With the experimental values of the parameters used it can be seen that the voltage across the contacts rises in a very short time to the minimum arcing voltage V_0, which is about 13·5 volts. If an arc then takes place, as will usually be the case, further calculations are necessary to determine the behaviour of the circuit. Fig. 1(*b*) shows the conditions in the circuit during an arc with voltage $v(t)$. The current flowing in the arc is I_v. Calculation of this current shows that with the range of values of the parameters used it differs little from the current I'_v in the equivalent circuit of the diagram given in Fig. 1(*c*). The inequality satisfied by I_v is given by

$$1{\cdot}1I'_v(t) - 0{\cdot}1I_B \leqslant I_v(t) \leqslant 0{\cdot}9(I_v(t) + 0{\cdot}1I_B \quad . \quad (2)$$

subject to $R \geqslant 0{\cdot}8\ \Omega$ and $0 \leqslant t \leqslant \frac{1}{2}L/R$, where t is measured from the commencement of arcing. It can be shown that the arc duration does not exceed $\frac{1}{2}L/R$ when $V = 6$ volts and $V_0 = 13{\cdot}5$ volts. The energy used in the arc is given by

$$W = \int_{t=0}^{t_A} v(t)I_v(t)dt$$

$$= \int_{I=I_B}^{I_A} v(t)I\left(\frac{dt}{dI}\right)dI \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where I_A is the extinction current of the arc. If the energy W' used in the arc of Fig. 1(*c*) is calculated assuming that $I_A \ll I_B$ and that $v(t) \geqslant V_0$ and using an equation of the type (3), then W' satisfies

$$(\tfrac{1}{2}LI_B^2) \leqslant W' \leqslant 1{\cdot}32(\tfrac{1}{2}LI_B^2) \quad . \quad . \quad . \quad . \quad (4)$$

If the inequality (2) is used it can be shown that, provided $v(t) \geqslant V_0 \geqslant 1{\cdot}2V + 1{\cdot}1RI_B = 2{\cdot}3V - 1{\cdot}1V_B$,

$$\left.\begin{aligned} &0{\cdot}73(\tfrac{1}{2}LI_B^2) \leqslant W \leqslant 1{\cdot}65(\tfrac{1}{2}LI_B^2) \text{ for } I_B \leqslant 5{\cdot}6 \text{ amperes} \\ \text{and}\quad &0{\cdot}95(\tfrac{1}{2}LI_B^2) \leqslant W \leqslant 1{\cdot}39(\tfrac{1}{2}LI_B^2) \text{ for } I_B \leqslant 1{\cdot}3 \text{ amperes} \end{aligned}\right\} \quad (5)$$

The calculations for eqn. (5) are simplified if it is assumed that the voltage $v(t)$ is constant but the inequalities are unchanged. Eqn. (5) gives the limits of the energy available in an arc at the contacts and hence the upper limit of the energy available for contact erosion.

The transfer was measured by optical estimation of the pip volumes as described previously.[1] This method was originated by Lander and Germer.[2] A photographic technique of measurement was used which consistently gave estimated values 2 lower than those obtained by using an eyepiece micrometer. For the results given below the error of measurement was most cases less than ±10%, but when the transfer was flat and spread over a wide area an attempt was made to allow for errors caused by the curvature of the wire and the different sh of the pip.

(3) RESULTS

The investigation was confined to the measurement of tran between platinum contacts breaking currents in a 6-volt circ The range of parameters used is given in Section 2. In subsequent discussion the circuit current will be taken as current I_B at break. It was observed that in all cases voltage, V_B, across the contacts at break was about 1·6 v and I_B can be obtained from I using this value of V_B. number of operations used for each measurement was $1{\cdot}5 \times$ It was necessary to use the same number of operations, since rate of transfer is not constant,[1] and complete investigation the variation of the rate of transfer with time would be leng A preliminary investigation at 1·3 amp showed that the tran with the new apparatus was the same as that observed with cam-operated contacts,[1] in spite of changes in rate of operat pressure and opening speed.

(3.1) Variation of Transfer with Inductance and Current

The variation of the average net cathode gain or loss operation is shown in Figs. 2 and 3 for $I_B = 1{\cdot}3$, 2·2, 2·9

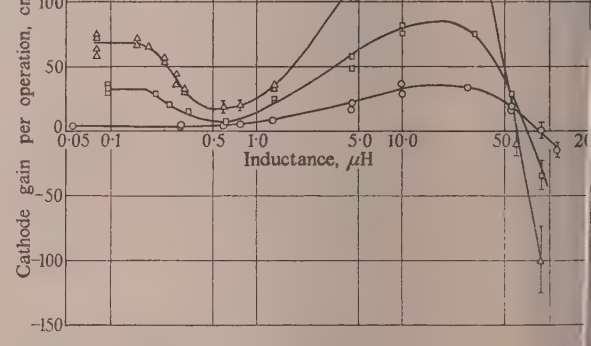


Fig. 2.—Variation of transfer with inductance and current ($1{\cdot}5 \times 10^5$ operations).
△ △ △ $I_B = 2{\cdot}9$ amp.
□ □ □ $I_B = 2{\cdot}2$ amp.
○ ○ ○ $I_B = 1{\cdot}3$ amp.

5·6 amp, and with inductances varying from the minimur to 117 μH. The gain or loss of metal from the anode wa measured, but where there was a well-defined cathode pip appeared to be only small total loss from anode and cat compared with the amount transferred. There was prol appreciable total loss for $I_B = 1{\cdot}3$, 2·2 and 2·9 amp wh was greater than 30 μH and for $I_B = 5{\cdot}6$ amp with $L >$ In these cases and for $I_B = 5{\cdot}6$ amp with $L > 0{\cdot}3\ \mu$H transfer was diffuse and not in sharp pip-and-crater form, s the estimation of volume gained or lost could be made between the limits shown on the curve. Nevertheless

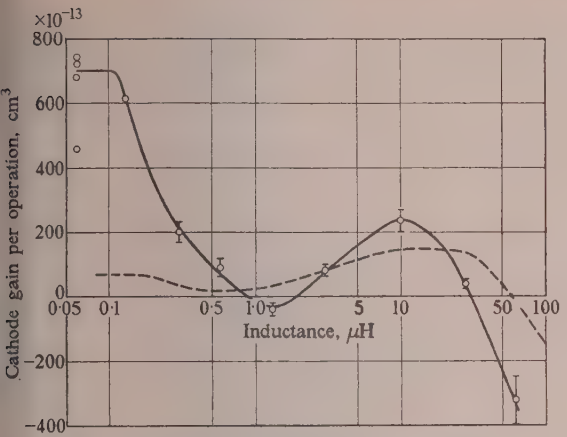


Fig. 3.—Variation of transfer with inductance for $I_B = 5{\cdot}5$ amp ($1{\cdot}5 \times 10^5$ operations).
○ ○ ○ $I_B = 5{\cdot}6$ amp.
— — — $I = 2{\cdot}9$ amp.

uracy was sufficient to establish the general magnitude and ection of transfer.

t can be seen that for all currents greater than about 2 amp net transfer varies with inductance in a complex manner g. 4). The values T_1, T_2, T_3, represent respectively the residual

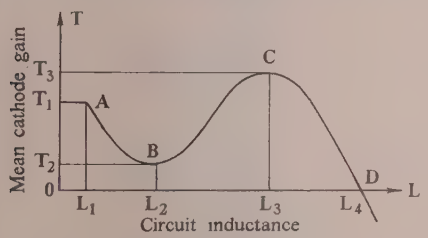


Fig. 4.—Variation of transfer with inductance.

ue, the first minimum, and the first maximum of the mean cathode gain. The values L_1, L_2, L_3, L_4 of inductance corrend to the first decrease, the first minimum, the first maximum l the final reversal of the transfer. The characteristics of the ves of Figs. 2 and 3 are summarized in Table 1.

Table 1

ARACTERISTICS OF CURVES FOR NET TRANSFER (FIGS. 2 AND 3)

urrent I_B	Inductance L				Mean net cathode gain per operation		
	L_1	L_2	L_3	L_4	T_1	T_2	T_3
amp	μH	μH	μH	μH	cm³ × 10⁻¹³	cm³ × 10⁻¹³	cm³ × 10⁻¹³
1·3	—	—	20	90	4	—	38
2·2	<0·2	0·5	15–20	70	33	12	85
2·9	0·18	0·6	15–20	60	67	19	153
5·6	<0·18	1·5	10	30	710	−50	230

he variation of net transfer with current for $L = L_0$, 0·6 μH 4·5 μH is shown in Fig. 5. The residual transfer is in all es a cathode gain and varies approximately as $I_B^{3\cdot4}$. The net nsfer for $L \geqslant 0{\cdot}6$ μH varies approximately as I_B^2 over certain ges of currents and inductances. These two relationships are

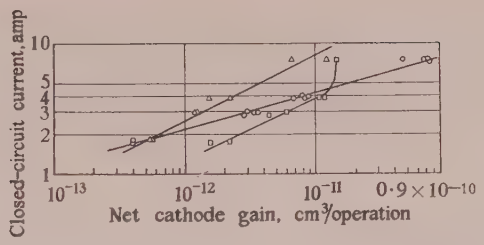


Fig. 5.—Variation of net transfer with current ($1{\cdot}5 \times 10^5$ operations).
△ △ △ $L = 0{\cdot}6$ μH.
□ □ □ $L = 4{\cdot}5$ μH.
○ ○ ○ $L = L_0$ (<0·1 μH).

purely empirical statements of the results obtained under the given experimental conditions.

Curves of the pip radius r (Fig. 6) and the ratio of pip height to pip radius (Fig. 7) show changes corresponding to the varia-

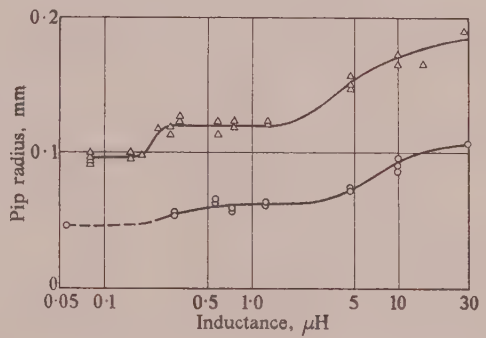


Fig. 6.—Variation of pip radius with inductance (after $1{\cdot}5 \times 10^5$ operations).
△ △ △ $I_B = 2{\cdot}9$ amp.
○ ○ ○ $I_B = 1{\cdot}3$ amp.

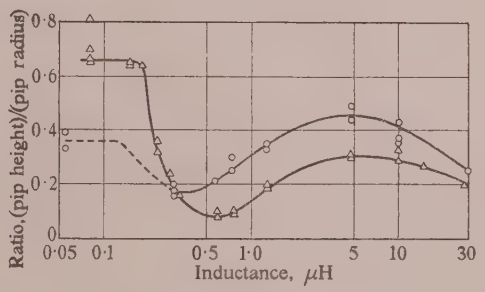


Fig. 7.—Variation of the ratio between pip height and pip radius (after $1{\cdot}5 \times 10^5$ operations).
△ △ △ $I_B = 2{\cdot}9$ amp.
○ ○ ○ $I_B = 1{\cdot}3$ amp.

tions in transfer. At minimum inductances the pips are tall and steep-sided, but when a small inductance of the order of 0·3 μH is introduced they become lower and wider. The base radius is very approximately proportional to I_B at all inductances. There is a close linear relationship at minimum inductance given by

$$r = 3{\cdot}6 \times 10^{-3} I_B \text{ centimetre for } 1{\cdot}3 \text{ amperes} \leqslant I_B \leqslant 5{\cdot}6 \text{ amperes} \quad . \quad . \quad (6)$$

The minimum value of r is probably determined, not by the current, but by the hardness of the platinum, and with a 30-g load the minimum would be about 3×10^{-3} cm.

A few tests were made with the contacts making as well as breaking the circuit, and when allowance had been made for bounce, which was slight, it appeared that the transfer was the same as that for break alone.

(3.2) Surface Structure of the Contacts and Surface Films

Examination of the surfaces of the contacts after a small number of operations shows that the changes of structure are similar to those described previously.[1] At the minimum inductances the anode and cathode are both finely pitted with craters of about equal diameter, in spite of the fact that the cathode gains material. Measurements show that the average values of the diameters, d, of these pits are proportional to the current, the relationship being given by the following equation:

$$\bar{d} = 2{\cdot}4 \times 10^{-4} I_B \text{ centimetre} = 1{\cdot}8 \times 10^{-4} I \text{ centimetre} \quad . \quad (7)$$

It is interesting to note that almost the same value of $\bar{d}$ was obtained for platinum over the same range of current by Pfann.[3] There is no marked change of surface structure in the region L_1, L_2 (Fig. 4), but, as the net cathode gain increases, the cathode structure becomes irregular while the individual pits in the anode increase steadily in diameter [see Fig. 6 of Reference (1)] even when L is greater than L_4.

It was observed that two types of film appeared in the neighbourhood of the contact areas. There was no evidence, however, to show that these films affected contact performance adversely. The first type was a smooth transparent film which showed as rings of interference colours round the contact area. In some cases the film had cracked off the surface, and in all cases there was a narrow ring adjacent to the contact area which was free from film. The films occurred on both contacts at all values of L greater than L_1 and increased in thickness with increasing values of L. The formation may be due to heating of the residue left on the surfaces after cleaning with trichloroethylene.

The second type of film appeared as a thick black ring round the contact area, graduating to a thin film of minute particles about 0·3 micron in diameter. These films obscured those of the first type and appeared on the anodes for inductances greater than 3 μH for all currents. There were traces of the same film on the cathodes at low inductances, but in most cases the area covered by the film remained small until L was greater than about 30 μH. At 5·6 amp, however, the dark ring appeared round the contact area of the cathode when L was 0·2 μH and increased slowly in area as L increased. It seems likely that the dark films consisted of finely divided platinum particles which were deposited on the surfaces either in the form of spray or condensation. The growth of the film appears to follow the trends of material transfer and to be associated with the contact which loses material.

(4) DISCUSSION

The results given in Section 3 show that the effect of inductance on transfer is more complex at higher than at lower currents. Thus the conclusions reached from measurements at 0·7 and 1·3 amp[1] do not apply to transfer at higher currents without modification. It is still possible, however, to deduce that there are two main types of transfer, namely bridge transfer and arc transfer. The arc transfer varies with inductance in a complex manner. For a given inductance the net transfer is the resultant of bridge transfer and up to three types of arc transfer. The following Sections contain an analysis of the observed transfer and a brief discussion of possible mechanisms involved.

(4.1) Bridge Transfer

There is a residual transfer, T_1, which does not depend on circuit inductance and which occurs in the form of very steep p on the cathode. The magnitude, Δ, of the transfer is given by empirical equation

$$\Delta(1{\cdot}5) = 1{\cdot}9 \times 10^{-13} \times I_B^{3{\cdot}4} \text{ cm}^3\text{/operation for}$$
$$1{\cdot}3 \text{ amperes} \leqslant I_B \leqslant 5{\cdot}6 \text{ amperes} \quad .$$

where $\Delta(1{\cdot}5)$ is the mean cathode gain for $1{\cdot}5 \times 10^5$ operatio The mechanism of transfer presumably is connected with formation of the molten bridge which is drawn between contacts and finally explodes when the hottest point reac boiling temperature.[1–4] According to the theory of the Thoms effect[5,6] the temperature gradients in the bridge just before boils are so high that the Thomson effect will cause a shift of hottest point, and the bridge will break asymmetrically. assumptions as to bridge dimensions and the value of the Tho son coefficient are made, it appears that the results of eqn. might be explained by the Thomson effect. At present, howev there is insufficient experimental evidence on which to base a pro

(4.2) Arc Transfer

It is evident that the transfer is dependent on the circ inductance when this is greater than a small value, L_1. It see that the only manner in which the inductive energy could ca transfer is by means of a discharge between the contacts. T conditions immediately after break are suitable for a transi discharge, since the bridge reaches boiling temperature bef separation[1,4] and the voltage across the contacts after bre reaches the minimum arcing voltage (about 13 volts) before current I_B has decreased appreciably. Examination of variation of surface structure and pip shape confirms hypothesis[1] that there is an arc between the contact immediat after break which causes metal transfer. Direct evidence of existence of an arc of very short duration between gold conta has been given by Lander.[11]

Recently it has been possible to observe arcs between platin contacts by using a high-speed oscillograph with a post-deflect acceleration tube. Records to be published at a later date sh that transient low-voltage discharges occur over the whole ra of measurement covered in this paper. It is shown in Sectio that the energy dissipated in an arc is of the order of $\frac{1}{2}I$ If the losses in the arc itself are small, as may be the case v very small gaps, most of the energy will be dissipated at electrodes and will cause evaporation and transfer. The electr which suffers greatest loss, and the efficiency of transfer, dep on the conditions in the gap. If there is sufficient energy avail the conditions may change once or twice before the arc fin ceases. This is the reason for the changes in the net tran which are observed as the inductance is increased. T changes and the possible mechanisms underlying them discussed below.

At the minimum inductances it appears that any disch which may occur at break does not affect the transfer. As inductance increases above a small value L_1 (see Table 1), t is a marked decrease in the steepness of the pips and an incr in base radius (Figs. 6 and 7). For currents greater than a 2 amp there is a decrease in net cathode gain which at 5·6 results in reversed transfer. This type of transfer is named convenience "reversed short-arc transfer." It has been obse by Allen[7] for currents of 10 and 20 amp but was not know exist for currents as low as 2 amp. A possible explanatio the mechanism is given by Allen in which he supposes th higher currents the metal vapour released into the gap v the bridge explodes is at several atmospheres pressure, so

e gap length at this pressure is equal to several electron mean ee paths and a long-arc type of discharge is set up in which the thode is bombarded by ions and loses material. Provided that ere is sufficient energy available, the arc will continue until the pour has escaped. If the gap length in terms of electron mean ee paths at atmospheric pressure is then sufficiently short the ual short-arc transfer occurs. This explains why the reversed ort-arc transfer is replaced by short-arc transfer as L increases. nce nothing is known about the surface tension of platinum boiling-point or the exact width of the contact gap immediately ter the explosion, it is not possible at present to develop the eory of reversed short-arc transfer.

A general explanation of short-arc transfer[1,8] in which the ode loses material is based on the supposition that the contact p after break is less than an electron mean free path, so that ost of the electrons reach the anode with 15 volts energy; this ergy is sufficient to evaporate the anode metal, some of which transferred to the cathode. A suggested mechanism by which e necessary electrons for the discharge are produced from the thode by the tunnel effect is given by Newton.[9] Some experiental work on this effect has been published recently.[10] The aracteristics of this transfer appear to be the same at all rrents, except that at higher currents it is complicated by the esence of reversed short-arc transfer.

The decrease in cathode gain and final reversal at high inctances is caused by the well-known long-arc erosion in which e cathode loses material by ion bombardment. The anode ins material, but there is probably an appreciable net loss. t this stage there is enough energy available for the arc to conue until the contacts have separated sufficiently for a long arc take place, so that the net transfer is the result of four superposed transfers of types: bridge, reversed short arc, short arc d long arc.

It was shown in Section 2 that the energy available in an arc pends on the arc voltage and the current I_B, and that for all lues of the circuit parameters which have been used it is of e order of the inductive energy $\frac{1}{2}LI_B^2$. A measure of the ficiency of total arc transfer is obtained by dividing the sum the absolute magnitudes of as many of the three types of arc ansfer as occur at a given inductance, L, by the total volume of atinum which could be evaporated by the energy $\frac{1}{2}LI_B^2$. Values the efficiency for a range of currents and inductances are ven in Table 2. The values may be up to 40% less or greater eqn. (5) is taken into account.

Table 2

EFFICIENCY OF ARC TRANSFER FOR FOUR VALUES OF I_B

L	Efficiency			
	$I_B = 1{\cdot}3$ amp	2·2 amp	2·9 amp	5·6 amp
μH	%	%	%	%
0·6	1·5	8·4	10·5	39·0
1·3	2·8	6·2	6·6	21·0
4·5	2·4	3·3	4·2	7·8
10·0	2·1	2·1	2·4	3·9
60·0	0·65	0·64	0·78	1·0
90·0	0·57	0·56	0·70	—

(4.3) Practical Considerations

A study of the results (Fig. 5) shows that improvement may be btained in contact life in practice if a small amount of inctance is included in the circuit. If this is done, the transfer spread over a wide area, and the tendency to form very steep ps is decreased; moreover, the net transfer can be reduced considerably without net loss. In the current range 1·3–5·6 amp the most suitable value of the effective circuit inductance is 0·6 μH. It would not be advantageous to add large inductances to obtain the small transfer observed in the region L_4, since this is not a stable position and the erosion is diffuse with considerable film growth and appreciable net loss.

(5) CONCLUSION

Experimental evidence shows that, with platinum contacts breaking a 6-volt circuit with break currents between 1·3 and 5·6 amp, the metal transfer between the contacts is the resultant of bridge transfer, which is independent of circuit inductance, and of arc transfer, which is dependent on inductance. The nature of the arc transfer is complex and up to three different types of arc transfer may be superimposed one on another if there is sufficient energy available. The types of transfer may be summarized as follows:

(*a*) Bridge transfer, which varies approximately as $I_B^{3\cdot4}$ and is independent of inductance. This transfer occurs as sharp spherical pips on the cathode for inductances less than about 0·2 μH but is masked by arc transfer for larger inductances.

(*b*) Reversed short-arc transfer, which occurs for $I_B > 2$ amp, and is the predominating arc transfer for inductances between about 0·2 μH and a value of the order of 1 μH (depending on the current). This transfer is from cathode to anode and can result in reversal of net transfer.

(*c*) Short-arc transfer, which predominates for inductances between 1 and 20 μH and is from anode to cathode.

(*d*) Long-arc transfer, which occurs for inductances greater than about 20 μH, is in the direction cathode to anode and leads to a final reversal in direction of the net transfer.

The last three types of transfer tend to be more diffuse, so that the pips are flatter than in residual transfer; moreover, the volume transferred does not vary so rapidly with I_B when arc transfer predominates. All the results apply only to transfer under the given conditions of contact pressure and opening speed. It is not likely, however, that the results are very sensitive to variations of pressure in the range of 10–100 g or of opening velocity when this is of the order of 1 cm/sec.

It appears that the bridge transfer is consistent with the Thomson effect, but the evidence is not conclusive. The reversed short-arc transfer may be due to excess pressure in the metal vapour between the separating contacts, but there is insufficient evidence at present to decide whether this is the case. The general principles of short-arc and long-arc transfer are understood, but no detailed theory of the mechanism has been developed. Further experimental work on other metals and a study of the arc characteristics would help towards the understanding of the phenomena observed. Preliminary work shows that the transfer between other metals in some cases differs quite widely from that between platinum contacts under similar conditions, so that the results for other metals cannot be deduced from those for platinum.

It should be possible to obtain considerably increased life with platinum contacts if the circuit is arranged so that the effective inductance is small (of the order of 0·6 μH).

(6) ACKNOWLEDGMENTS

The author wishes to thank the Director of the British Electrical and Allied Industries Research Association for permission to publish the paper and to express appreciation of the interest shown by Mr. W. Nethercot during the course of the work.

(7) REFERENCES

(1) WARHAM, J.: "The Effect of Inductance on Fine Transfer between Platinum Contacts," *Proceedings I.E.E.*, Paper No. 1504, July, 1953 (**100**, Part 1, p. 163).

(2) LANDER, J. J., and GERMER, L. H.: "Bridge Erosion of Electrical Contacts, Part 1," *Journal of Applied Physics*, 1948, **19**, p. 910.

(3) PFANN, W. G.: "Bridge Erosion of Electrical Contacts and its Prevention," *Transactions of the American I.E.E.*, 1948, **67**, p. 1528.

(4) HOLM, R.: "Electric Contacts" (Gebers Forlag, Stockholm, 1946).

(5) LANDER, J. J.: "Measurements of Thomson Coefficients for Metals at High Temperatures," *Physical Review*, 1948, **74**, p. 479.

(6) DAVIDSON, P. M.: "The Theory of the Thomson Effect in Electrical Contacts," *Proceedings I.E.E.*, Paper No. 871, November, 1949 (**96**, Part 1, p. 295).

(7) ALLEN, A. L.: "Long-Life Contacts for Unidirectiona Currents of 1–20 Amperes," *ibid.*, Paper No. 1506, July 1953 (**100**, Part I, p. 158).

(8) GERMER, L. H., and HAWORTH, F. E.: "Erosion of Electrica Contacts on Make," *Journal of Applied Physics*, 1949, 20 p. 1085.

(9) NEWTON, R. R.: "Ejection of Electrons by Ions at Hig Fields," *Physical Review*, 1948, **73**, p. 1122.

(10) LLEWELLYN JONES, F., and MORGAN, C. G.: "Surface Film and Field Emission of Electrons," *Proceedings of th Royal Society*, A, 1953, **218**, p. 88.

(11) LANDER, J. J.: "Use of an Electron Diffraction Camera as a Ultraoscilloscope, and Suggested Application to Contac Erosion," *Journal of Applied Physics*, 1949, **20**, p. 108

621.317.79: 621.372.5: 519.241.1: 518.5

Monograph No. 104 R
Aug. 1954

A NEW METHOD OF DETERMINING CORRELATION FUNCTIONS OF STATIONARY TIME SERIES

By D. G. LAMPARD, M.Sc.

(*The paper was first received 6th March, and in revised form 29th April*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in August*, 1954.)

SUMMARY

The paper presents what is believed to be a new approach to the problem of determining correlation functions of stationary time series. The method is based on an expansion of the correlation function in a suitable orthonormal system. The coefficients in this expansion are shown to be simply related to the convolution integrals which arise when the given time series are applied to linear filters whose impulse responses are members of this orthonormal system. A very simple scheme for measuring these coefficients is proposed. It is further shown how a complex filter, whose impulse response is of the same shape as the desired correlation function, can be built up in a systematic way when these coefficients have been determined. Two examples of filter systems whose impulse responses are members of a Laguerre system are presented, and finally a description of a practical auto-correlator built around one of these filters, and some results obtained with it, are given.

(1) INTRODUCTION

The importance of statistical concepts and methods for handling problems in such diverse fields as radio noise, turbulence, tidal motion, astronomy, ionospheric physics, etc., is now firmly established.

A parameter which occurs frequently in the discussion of many of these problems is the temporal "correlation function" of two stationary time series $x_1(t)$ and $x_2(t)$ which is usually defined (Reference 2, p. 5) as

$$\psi_{12}(\tau) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{+T} x_1(t)x_2(t+\tau)dt \quad . \quad . \quad . \quad (1)$$

or

$$\psi_{12}(\tau) = \overline{x_1(t)x_2(t+\tau)} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where the bar denotes time averaging in the sense specified in eqn. (1).

Because of the importance of this function it is natural that devices should be required, which, when presented with the time series $x_1(t)$ and $x_2(t)$, should carry out the operations implied by eqn. (1), thus enabling the function to be determined experimentally.

It is seen from eqns. (1) and (2) that the essential feature of this definition is the temporal delay τ between the $x_1(t)$ and $x_2(t+\tau)$. When the $x_1(t)$ and $x_2(t)$ are fluctuating voltages or currents the provision of this "pure" delay presents rather a problem.

For very short delays (less than 100 microsec, say) artificial lines have been constructed to give the required delay without appreciable distortion of the waveform, while methods which have been used satisfactorily for longer delays include tape or film recording followed by playback with displaced reproducing heads, and pulse sampling methods.

In the paper a method which is believed to be a new approach to the problem of determining correlation functions will be presented. This method has the advantage that a "pure" delay is not required.

(2) THEORY

(2.1) Expansion in Orthogonal Functions—General Case

Suppose that any correlation function* $\psi(\tau)$ can be expanded in a series of orthogonal functions (Reference 3, p. 22, and Reference 4, p. 49), thus

$$\psi(\tau) = \sum_{n=0}^{\infty} a_n\theta_n(\tau)[\omega(\tau)]^{\gamma},\ 0 \leqslant \tau < \infty \quad . \quad . \quad (3)$$

where the $\theta_n(\tau)$ are polynomials which form an orthonormal set with respect to the weight function $\omega(\tau)$ in the range $0 \leqslant \tau < \infty$. Thus

$$\int_0^{\infty} \theta_n(\tau)\theta_m(\tau)\omega(\tau)d\tau = \left.\begin{matrix} 0,\ m \neq n \\ 1,\ m = n \end{matrix}\right\} \quad . \quad . \quad . \quad (4)$$

Then the coefficients a_n in the series expansion shown in eqn. (3) can be found in the usual way and are given by

$$a_n = \int_0^{\infty} \psi(\tau)\theta_n(\tau)[\omega(\tau)]^{1-\gamma}d\tau \quad . \quad . \quad . \quad . \quad (5)$$

Also, it may be shown (Reference 3, p. 24) that if

$$\psi_N(\tau) = \sum_{n=0}^{N} b_n\theta_n(\tau)[\omega(\tau)]^{\gamma} \quad . \quad . \quad . \quad . \quad (6)$$

then

$$\int_0^{\infty} [\psi(\tau) - \psi_N(\tau)]^2[\omega(\tau)]^{1-2\gamma}d\tau \quad . \quad . \quad . \quad (7)$$

is a minimum if, and only if,

$$b_n = a_n \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

In other words if only a finite number, $N+1$, of terms are going to be retained in the series expansion of $\psi(\tau)$, then the approximation will be best in a suitably weighted "least squares" sense if the coefficients of the retained terms have the same values as those in the infinite series.

(2.2) The Symmetric Case $\gamma = \frac{1}{2}$

If $\gamma = \frac{1}{2}$, eqns. (3) and (5) may be written in the "symmetric" form

$$\psi(\tau) = \sum_{n=0}^{\infty} C_n\phi_n(\tau),\ 0 \leqslant \tau < \infty$$

$$C_n = \int_0^{\infty} \psi(\tau)\phi_n(\tau)d\tau \quad . \quad . \quad . \quad . \quad (10)$$

where

$$\phi_n(\tau) = \theta_n(\tau)[\omega(\tau)]^{\frac{1}{2}} \quad . \quad . \quad . \quad . \quad . \quad (11)$$

* The subscripts will be dropped when not necessary for the argument.

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Lampard, who was formerly in the Division of Electrotechnology of the Commonwealth Scientific and Industrial Research Organization, Australia, is at Corpus Christi College, Cambridge.

It is easily seen from eqn. (7) that in this case the partial sums are best approximations in a "least square" sense with uniform weighting.

(2.3) Linear Networks and "Orthogonal" Filters

Let us now consider the system of Fig. 1. We suppose that the linear network has an impulse response $h_n(t)$. Then it

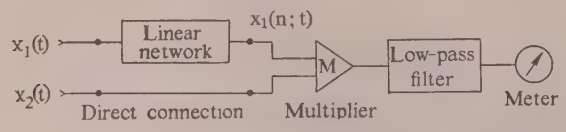


Fig. 1.—Basic measuring system.

follows from the superposition theorem that the filter output $X_1(n; t)$ is given by

$$X_1(n; t) = \int_0^\infty x_1(t-u)h_n(u)du \quad . \quad . \quad . \quad (12)$$

The multiplier output is then

$$X_1(n; t)x_2(t) = \int_0^\infty x_1(t-u)x_2(t)h_n(u)du \quad . \quad . \quad (13)$$

and the mean value* of eqn. (13) is

$$\overline{X_1(n; t)x_2(t)} = \int_0^\infty \overline{x_1(t-u)x_2(t)}h_n(u)du \quad . \quad . \quad (14)$$

$$= \int_0^\infty \psi_{12}(u)h_n(u)du \quad . \quad . \quad . \quad . \quad (15)$$

In passing from eqn. (14) to eqn. (15), use has been made of the stationary nature of the $x_1(t)$ and $x_2(t)$, and also of the definition in eqn. (2). Now comparing eqn. (15) with eqn. (5) it is readily seen that the coefficient a_n is identical with the average $\overline{X_1(n; t)x_2(t)}$ provided that the filter impulse response is so chosen that

$$h_n(t) = \theta_n(t)[\omega(t)]^{1-\gamma} \quad . \quad . \quad . \quad . \quad (16)$$

Thus we are led to the concept of a set of "orthogonal filters" whose impulse responses satisfy eqn. (16). If such a set of filters could be realized the coefficients a_n would be obtained simply by switching in each of the filters in turn and determining the corresponding averages $\overline{X_1(n; t)x_2(t)}$.

Naturally, in any practical system only a finite number, say $(N+1)$, of filters would be used.

(2.4) Resynthesis of the Correlation Function

We see that eqn. (3) can be written

$$\psi(\tau) = [\omega(\tau)]^{2\gamma-1} \sum_{n=0}^{\infty} a_n\theta_n(\tau)[\omega(\tau)]^{1-\gamma}, \; 0 \leqslant \tau < \infty \quad . \quad (17)$$

or

$$\psi(t) = [\omega(t)]^{2\gamma-1} \sum_{n=0}^{\infty} a_n h_n(t), \; 0 \leqslant t < \infty \quad . \quad . \quad (18)$$

where, in passing from eqn. (17) to eqn. (18), use has been made of eqn. (16) coupled with a change of variable from τ to t.

* In practice this mean value would be determined by passing the multiplier output through a suitable low-pass filter or integrating device whose time-constant was sufficiently long to reduce the residual fluctuations about the mean to an acceptable value. Expressions for the r.m.s. value of this fluctuation have been given for a number of averaging devices by Davenport *et al.*[5]

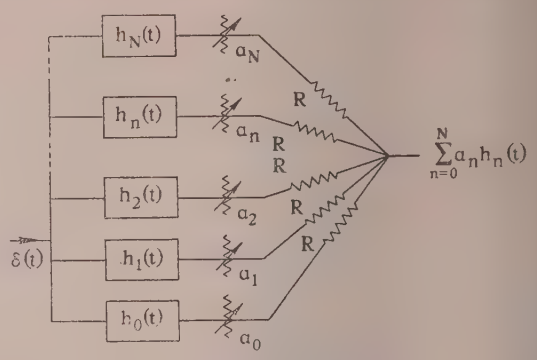


Fig. 2.—Basic resynthesis system.

Now consider the system of Fig. 2. Here each filter is follow by an adjustable attenuator and the outputs of the attenuato are summed in a linear adding network. The attenuators a set so that their transmissions are proportional to the previous measured coefficients† a_n.

Clearly, then, if a unit impulse $\delta(t)$ is applied at the input the system, the output of the adding network is just the sur mation term of eqn. (18) approximated by taking only t first $(N+1)$ terms.

Thus we may write as an approximation to $\psi(t)$,

$$\psi_N(t) = [\omega(t)]^{2\gamma-1} \sum_{n=0}^{N} a_n h_n(t), \; 0 \leqslant t < \infty \quad . \quad (1$$

In particular, if $\gamma = \frac{1}{2}$, we have

$$\psi_N(t) = \sum_{n=0}^{N} a_n h_n(t), \; 0 \leqslant t < \infty \quad . \quad . \quad . \quad (2$$

when now, using the notation of eqn. (8), we must have

$$h_n(t) = \phi_n(t) \quad . \quad . \quad . \quad . \quad . \quad . \quad (2$$

Thus in this case it is seen that the output of the adding n work could be applied direct to the plates of a cathode-r oscillograph; the approximate correlation function would th be seen on the screen as a transient following the applicati of the input impulse. A convenient system results if in fact t pulsing can be carried out repetitively at intervals that are lo compared with the filter time-constants, and yet sufficiently f quently to give a steady trace on the cathode-ray-oscillogra screen. This can always be arranged in practice, as the res thesis filters, while being of the same form as those used for determination of the coefficients, can be built on a different ti scale.

It will be noticed from eqn. (18) that these methods give correlation function $\psi(t)$ only for $0 \leqslant t < \infty$, but in gene negative values of the delay parameter τ are also requir However, it follows from the definition in eqn. (2) that

$$\psi_{12}(-\tau) = \psi_{21}(\tau) \quad . \quad . \quad . \quad . \quad . \quad (2$$

so that it is necessary only to repeat the procedure with "linear network" and the "direct connection," in the system Fig. 1, interchanged, to cater for all values of τ. The diffic does not arise, of course, in the case of auto-correlation.

(2.5) Two Realizable Systems based on the Weight Func $\omega(t) = \alpha\varepsilon^{-\alpha t}$

The applicability of all the preceding theory depends realizing a set of filters whose impulse responses satisfy eqn. (

† In general some of the a_n would be negative, so it is assumed that an ideal r reversal is provided where necessary (see Section 3.1).

The polynomials $\theta_n(t)$ and hence the impulse responses $h_n(t)$ depend only on the weight function $\omega(t)$ and the parameter γ.

In this Section we shall choose $\omega(t) = \alpha\varepsilon^{-\alpha t}$ and show that this choice leads to realizable filters of a simple type. Two such filters, corresponding to $\gamma = 0$ and $\gamma = \frac{1}{2}$ respectively, will be described.

The polynomials associated with the weight function $\alpha\varepsilon^{-\alpha t}$ in the range zero to infinity are just those of Laguerre, whose series representation, in a form convenient for our purposes, is*

$$L_n(\alpha t) = \sum_{s=0}^{n} \frac{n!}{(n-s)!} \frac{(-\alpha t)^s}{s!s!} \quad . \quad . \quad . \quad (23)$$

The relation corresponding to eqn. (4) is

$$\int_0^\infty \alpha\varepsilon^{-\alpha t} L_n(\alpha t) L_m(\alpha t) dt = \left.\begin{matrix} 0 & m \neq n \\ 1 & m = n \end{matrix}\right\} \quad . \quad . \quad (24)$$

As a further preliminary we now introduce the Laplace transform of a function $f(t)$ defined (Reference 6, p. 4) by the relation

$$\mathscr{L}_p[f(t)] = \int_0^\infty f(t)\varepsilon^{-pt} dt \quad . \quad . \quad . \quad . \quad (25)$$

By using the series expansion of eqn. (23), interchanging the order of summation and integration, and noting that the resulting series is just a binomial expansion, it is easy to show that

$$\mathscr{L}_p[\alpha\varepsilon^{-\alpha t} L_n(\alpha t)] = \frac{\alpha}{p+\alpha}\left(\frac{p}{p+\alpha}\right)^n \quad . \quad . \quad . \quad (26)$$

$$\mathscr{L}_p\left[\frac{\alpha}{2}\varepsilon^{-\frac{\alpha}{2}t} L_n(\alpha t)\right] = \frac{\alpha/2}{p+\alpha/2}\left(\frac{p-\alpha/2}{p+\alpha/2}\right)^n \quad . \quad . \quad (27)$$

Now consider the filter of Fig. 3.

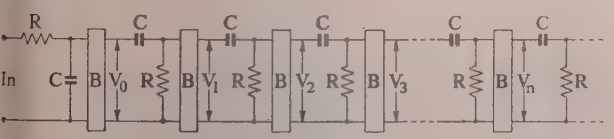


Fig. 3.—Laguerre function filter: $\gamma = 0$.

Here B is a buffer stage which has, ideally, an infinite input resistance, zero output resistance and unity gain. However, non-ideal values of these parameters are easily allowed for, while in practice a cathode follower is virtually "ideal."

It is easy to see that the recurrence relation

$$\mathscr{L}_p[V_{n+1}(t)] = \frac{p}{p+\alpha}\mathscr{L}_p[V_n(t)] \quad . \quad . \quad . \quad (28)$$

is satisfied for $0 \leqslant n < N$ where $\alpha = \frac{1}{RC}$.

* Using the Pochammer–Barnes (Reference 3, p. 99) notation for the confluent hypergeometric function, eqn. (23) may be written

$$L_n(\alpha t) = {}_1F_1[-n;\ 1;\ \alpha t]$$

Thus it follows that the impulse response $h_n(t)$, which is just the $V_n(t)$ obtained when $\delta(t)$ is applied at the input, satisfies the equation

$$\mathscr{L}_p[h_n(t)] = \frac{\alpha}{p+\alpha}\left(\frac{p}{p+\alpha}\right)^n \quad . \quad . \quad . \quad . \quad (29)$$

On comparing eqn. (29) with eqn. (26), it follows from Lerch's theorem (Reference 6, p. 345) on the uniqueness of Laplace transforms that

$$h_n(t) = \alpha\varepsilon^{-\alpha t} L_n(\alpha t) \quad . \quad . \quad . \quad . \quad . \quad (30)$$

which is of the form shown in eqn. (16) with $\gamma = 0$ for the given weight function.

Another filter based on the same weight function is the lattice structure of Fig. 4.

In this case it is easy to show that the recurrence relation corresponding to eqn. (28) is now

$$\mathscr{L}_p[V_{n+1}(t)] = \left(\frac{p-\alpha/2}{p+\alpha/2}\right)\mathscr{L}_p[V_n(t)] \quad . \quad . \quad (31)$$

where $\frac{\alpha}{2} = \frac{1}{RC}$.

Thus the impulse response satisfies the equation

$$\mathscr{L}_p[h_n(t)] = \frac{\alpha/2}{p+\alpha/2}\left(\frac{p-\alpha/2}{p+\alpha/2}\right)^n \quad . \quad . \quad . \quad (32)$$

and a comparison of eqns. (32) and (27) now gives

$$h_n(t) = \frac{\alpha}{2}\epsilon^{-\frac{\alpha}{2}t} L_n(\alpha t) \quad . \quad . \quad . \quad . \quad . \quad (33)$$

which, except for a constant factor, is again of the form shown in eqn. (16) but with $\gamma = \frac{1}{2}$. Thus this filter is the "symmetric" case.

We note† that this last filter is related to one devised by Lee.[7] Lossless inductances are required in his filter, however, and would lead to difficulties especially at low frequencies. Lee's filters were developed to provide a systematic way of realizing arbitrary filter impulse responses, and it is obvious that the filters described above could be used in the same way.

We note finally in this Section that the use of an exponential weight function is sensible from the statistical point of view as the correlation function of a simple Markovian process is exponential and hence is represented by the first term only of the series expansion in the Laguerre system, provided that the parameter α (i.e. the time-constant) in the weight function is chosen appropriately.

In a recent paper Reich and Swerling[8] have also proposed that an expansion of auto-correlation functions in a Laguerre system might be convenient in connection with some detectability problems.

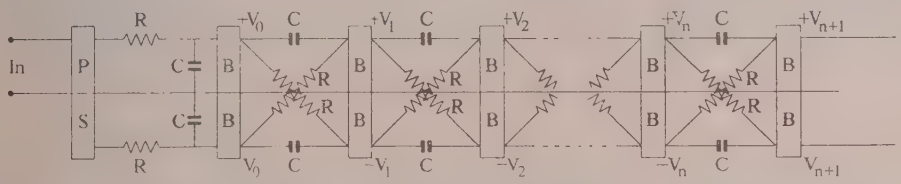


Fig. 4.—Laguerre function filter: $\gamma = \frac{1}{2}$.
PS—phase splitter.

† The author is indebted to Mr. Newstead, of the University of Tasmania, for bringing this reference to his attention.

(3) A PRACTICAL CORRELATOR

An auto-correlator based on an 11-stage ($N = 10$) filter of the type shown in Fig. 3 has been built. Although not quite as convenient in some ways as the "symmetric" filter, this filter is simpler to construct and in any case adequately demonstrates the main features of the method in the paper. It will be seen from eqn. (7) that for this filter the integral

$$\int_0^\infty [\psi(\tau) - \psi_N(\tau)]^2 \varepsilon^{-\alpha\tau} d\tau \quad . \quad . \quad . \quad . \quad (34)$$

is minimized, and in practice this may be rather more desirable than the uniform weighting obtained with the "symmetric" filter.

(3.1) Circuit Details

The complete circuit of the filter is shown in Fig. 5. It will be seen that the electronic circuit technique is very simple:

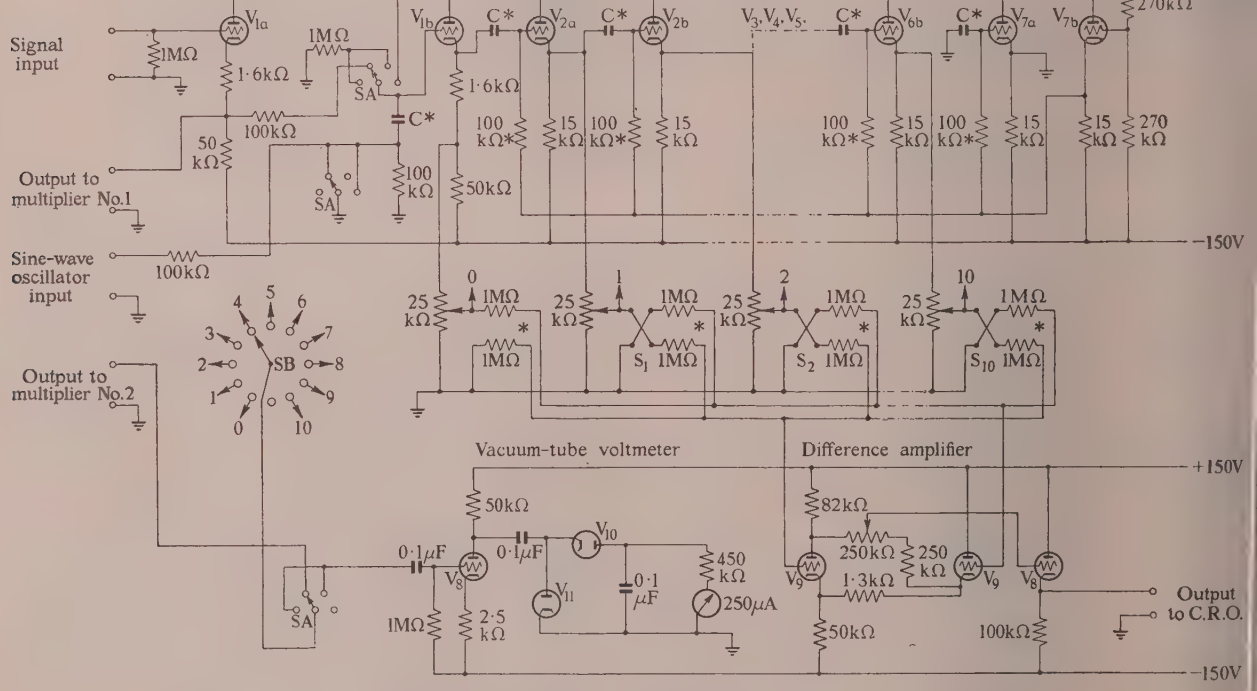


Fig. 5.—Laguerre function filter: detailed circuit.

Switch SA
(1) Set potentiometers.
(2) Measure.
(3) Reset potentiometer.
(4) Pulse.

V_1–V_9: 6SN7.
V_{10}–V_{11}: EA50.
C: Condensers (mounted in separate plug-in unit).

* denotes "to be matched carefully."

with the exception of the peak-to-peak diode voltmeter and the difference amplifier, it consists entirely of cathode followers. By means of V7 and its associated components, all the cathodes of valves V2–V6 are arranged to be normally at earth potential, and hence the d.c. levels of the outputs to the multiplier are independent of the "coefficient" potentiometer settings. The condensers, C, are all mounted in a single plug-in unit so that the basic time-constant of the filter can be changed at will.

The operation of the correlator will be most conveniently described in four separate stages corresponding to the setting of switch SA.

Stage 1.—With switch SA in position 1, a sine wave, whos frequency is high (say 25 kc/s when $\alpha = 1\,000$) compared to th reciprocal of the basic filter time-constant, is fed into the gri circuit of the buffer stage V1b. By means of switch SB th vacuum-tube voltmeter is connected to the output of each filte stage in turn and the "coefficient" potentiometers are adjuste so that the outputs of all stages are identical, thus compensatin for the cumulative effect of the small cathode-follower losses.

Stage 2.—Switch SA is next placed in position 2 which cor nects the signal whose auto-correlation is to be determined t the input of the filter. One input to the multiplier is permanentl connected to this signal input, while, by means of switch SE the other multiplier input is connected to the output of eac stage of the filter in turn and the corresponding mean product are determined.

The multiplication and averaging is performed by a feedbac

dynamometer system followed by a Miller integrator whose tim constant is variable from 1 to 1 000 sec. The mean produc are read direct on a voltmeter connected to the output of t integrator. The accuracy of the multiplier is about 1%, and frequency response is essentially flat from direct current up about 15 kc/s. A description of the multiplier and integrat will be published elsewhere.

Stage 3.—With switch SA in position 3 the high-frequer sine wave is again connected to the buffer stage V1b. Using t

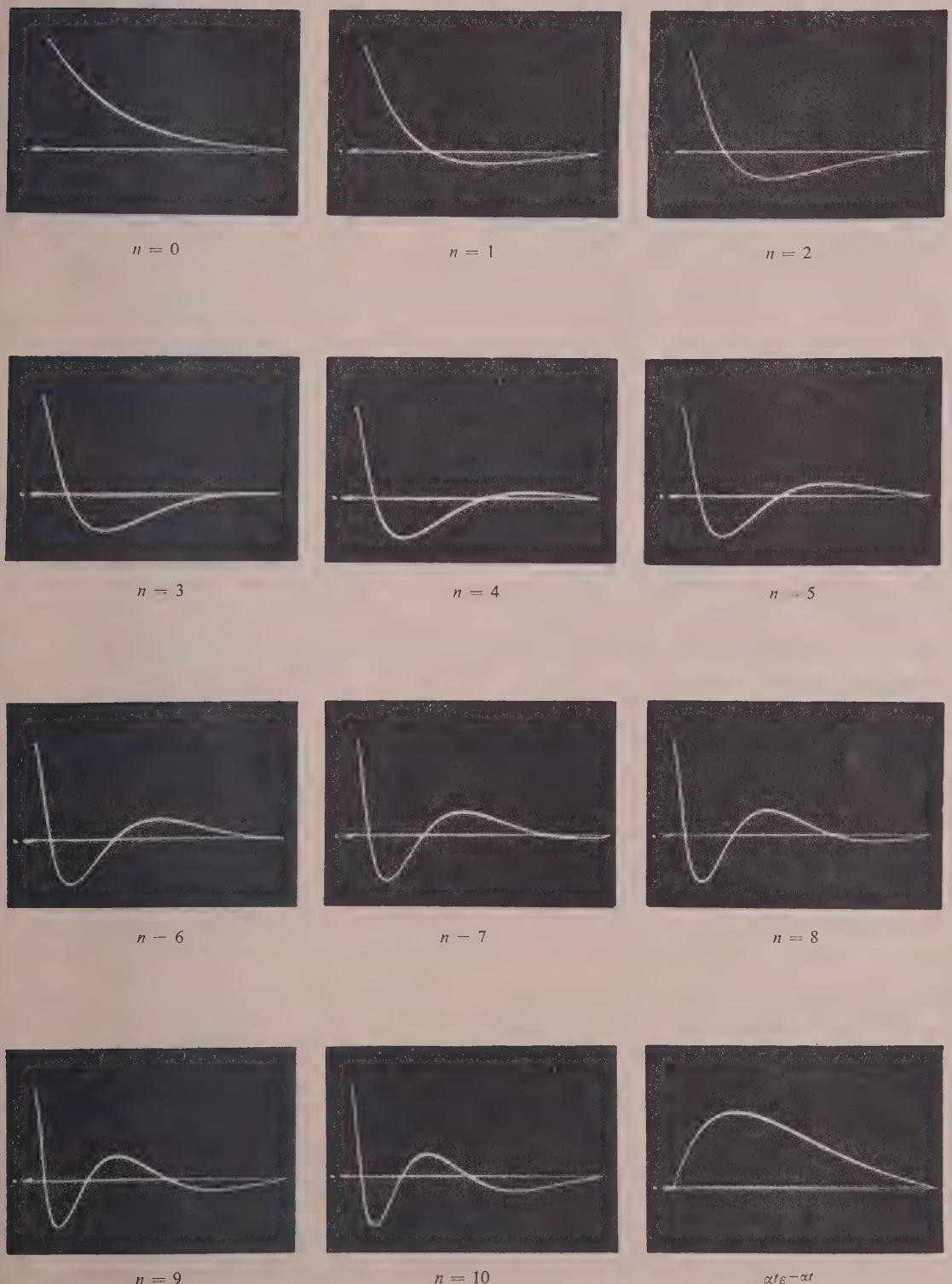


Fig. 6.—The Laguerre functions $\alpha\varepsilon^{-\alpha t}L_n(\alpha t)$.

vacuum-tube voltmeter and switch SB the "coefficient" potentiometers are now readjusted so that the output from each stage of the filter is proportional to the corresponding mean product determined in the previous step. In general, some of these mean products will be negative. To cope with these in the resynthesis process, two summing networks are provided, their outputs being connected to the two inputs of a "perfect" difference amplifier. By means of the switches S1–S10 the output of any given stage of the filter can be connected to either one of these summing networks, so that its contribution to the difference amplifier output may be made positive or negative.

Stage 4.—Finally by placing switch SA in position 4 a pulse generator which is triggered by the cathode-ray-oscillograph time-base flyback pulse is connected to the filter input, and the output from the difference amplifier is applied to the vertical plates of the oscillograph. With this filter the function $\alpha\varepsilon^{-\alpha t}\psi_N(t)$ is then obtained on the screen. A simple multiplication by $\epsilon^{\alpha t}/\alpha$ then yields the desired approximate auto-correlation function, $\psi_N(t)$. This last multiplication would have been avoided, of course, if the "symmetric" filter had been used instead. The time scale of the oscillograph trace is easily calibrated by comparison with the trace of $\alpha\varepsilon^{-\alpha t}L_n(\alpha t)$, whose zero occurs at $\alpha t = 1$.

(3.2) Results

The impulse responses $\alpha\varepsilon^{-\alpha t}L_n(\alpha t)$ were photographed direct from the oscillograph screen and are shown in Fig. 6. Because of the heavy damping by the exponential factor and the time scale used, only the first few zeros are apparent.

To demonstrate the difference-amplifier system, mentioned in the previous Section, the function

$$\alpha(\alpha t)\varepsilon^{-\alpha t} = [\alpha\varepsilon^{-\alpha t}L_0(\alpha t)] - [\alpha\varepsilon^{-\alpha t}L_1(\alpha t)] \quad . \quad . \quad (35)$$

is also shown in Fig. 6. The possibilities of the system for generating arbitrary transients are obvious.

As a test on the performance of the whole system as an auto-correlator, a sine-wave input was used. The auto-correlation function of a sine wave* of peak amplitude V_0 and frequency $\omega/2\pi$ is just

$$\psi(\tau) = \frac{V_0^2}{2}\cos\omega\tau \quad . \quad . \quad . \quad . \quad . \quad (36)$$

The coefficients a_n are then given by

$$a_n = \frac{V_0^2}{2}\int_0^\infty \alpha\varepsilon^{-\alpha\tau}L_n(\alpha\tau)\cos\omega\tau d\tau \quad . \quad . \quad . \quad (37)$$

By writing the cosine as the sum of exponentials, using the series of eqn. (23), followed by term-wise integration, it is easy to show that

$$a_n = \frac{V_0^2}{2}\,\frac{\alpha}{\sqrt{(\alpha^2+\omega^2)}}\left[\frac{\omega}{\sqrt{(\alpha^2+\omega^2)}}\right]^n \sin(n+1)\theta \quad . \quad (38)$$

where $\tan\theta = \dfrac{\alpha}{\omega}$

$$= \frac{V_0^2}{2}\,\frac{\alpha\omega}{\alpha^2+\omega^2}\left[\frac{\omega}{\sqrt{(\alpha^2+\omega^2)}}\right]^n U_n\left[\frac{\omega}{\sqrt{(\alpha^2+\omega^2)}}\right] \quad (39)$$

where $U_n(x)$ is the Chebȳshev polynomial (Reference 3, p. 3) of the second kind.

* The frequency of the sine wave used was 152 c/s, and the basic time-constant of the filter was such that $\alpha = 843\ \text{sec}^{-1}$.

It follows from eqn. (40) that†

$$\frac{a_n}{a_1} = \frac{x^{n-2}U_n(x)}{2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

where $x = \dfrac{\omega}{\sqrt{(\alpha^2+\omega^2)}}$.

This ratio, as given by eqn. (40), is plotted in Fig. 7. Also shown are the corresponding ratios determined experimentally

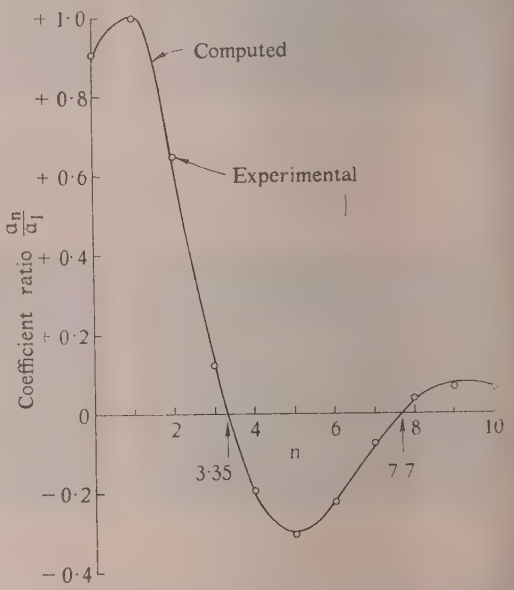


Fig. 7.—Expansion coefficients for cosine correlation function in Laguerre system: $\gamma = 0$.

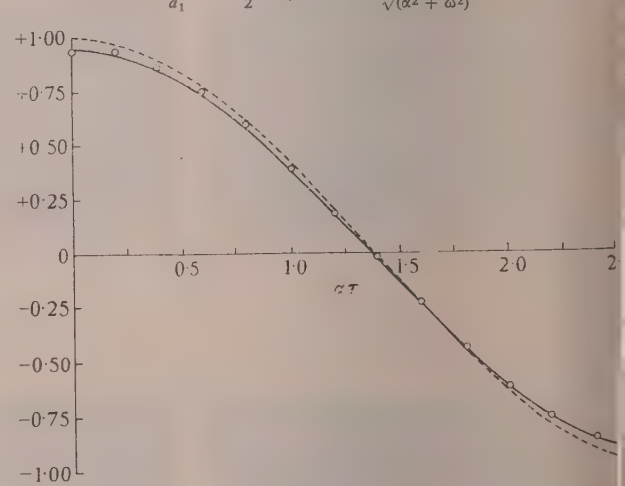


Fig. 8.—The approximation of a cosine correlation function by the partial sum of its series expansion in Laguerre polynomials.

——— $\cos\left(\frac{\omega}{\alpha}\alpha\tau\right)$

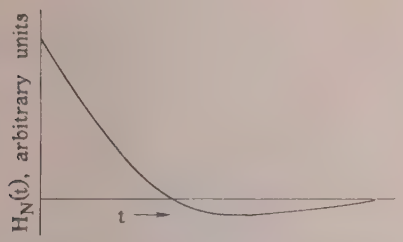
——— $\sum_{n=0}^{10} a_n L_n(\alpha t)$ with $a_n = \int_0^\infty \alpha\varepsilon^{-\alpha\tau}L_n(\alpha\tau)\cos\omega\tau d\tau$

○ ○ ○ ○ Points giving $\sum_{n=0}^{10} a'_n L_n(\alpha\tau)$, with a'_n determined experimentally.

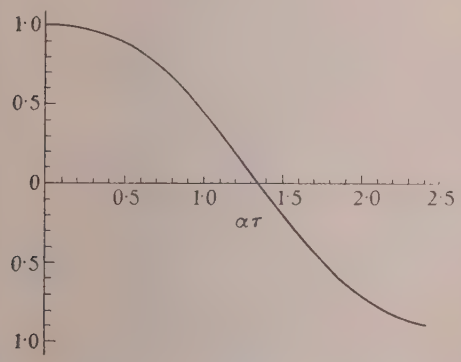
† It is readily seen that the coefficient a_0 depends on the d.c. level (mean) of the input whose auto-correlation function is being determined. For this reason the coefficient ratios are taken with respect to a_1 rather than a_0.

n Fig. 8 the first half-cycle of the correlation function, eqn. (36), s shown. In the same figure a curve showing the sum of the irst 11 terms in its Laguerre polynomial expansion is plotted, nd points representing this same sum—but based on measured ather than calculated coefficients—are also given for comparison.

We note here that several examples of expressions of simple unctions in a Laguerre system, and their approximation by a nite number of terms, have been given in a recent paper by Vard (Reference 9).



ig. 9.—Oscillographic trace of impulse response $H_N(t)$ of composite filter.



ig. 10.—Normalized cosine correlation function recovered by exponential weighting of impulse response of composite filter (see Fig. 9).

Finally the correlation function as determined by the re-synthesis process is shown in Fig. 10, and it is readily seen to be a good approximation to the true correlation function of eqn. (36).

(4) ACKNOWLEDGMENTS

The author is grateful to Miss L. J. Slater of the Mathematical Laboratory, Cambridge, for her kindness in supplying a special table of Laguerre polynomials, and to Mr. J. F. Barrett and Dr. K. F. Sander for reading the text and suggesting several improvements. This work was carried out while the author was the holder of a C.S.I.R.O. (Australia) Studentship and was working at the Engineering Laboratory, Cambridge, under the supervision of Dr. K. F. Sander.

(5) BIBLIOGRAPHY

(1) Stumpers, F. L.: "A Bibliography of Information Theory" (R.L.E. M.I.T., Cambridge, Massachusetts, 1953).

(2) Wiener, Norbert: "Extrapolation, Interpolation and Smoothing of Stationary Time Series" (Technology Press of M.I.T. and John Wiley and Sons).

(3) Szegö, G.: "Orthogonal Polynomials" [American Mathematical Society Colloquium Publications. Vol. 23 (American Mathematical Society, 1939)].

(4) Courant, R., and Hilbert, D.: "Methods of Mathematical Physics" (Interscience Publishers Incorporated, 1953), Vol. 1.

(5) Davenport, W. B., Johnson, R. A., and Middleton, D.: "Statistical Errors in Measurements on Random Time Functions," *Journal of Applied Physics*, 1952, **23**, p. 377.

(6) Carslaw, H. S., and Jaeger, J. C.: "Operational Methods in Applied Mathematics" (Oxford University Press, 1948).

(7) Lee, Y. W.: "Synthesis of Electric Networks by means of the Fourier Transforms of Laguerre Functions," *Journal of Mathematics and Physics*, 1931–32, **11**, p. 83.

(8) Reich, E., and Swerling, P.: "The Detection of a Sine Wave in Gaussian Noise," *Journal of Applied Physics*, 1953, **24**, p. 289.

(9) Ward, E. E.: "The Calculation of Transients in Dynamical Systems," *Proceedings of the Cambridge Philosophical Society*, **50**, Part 1, p. 49.

621.3.015.5:537.525

Monograph No. 105
Aug. 195

FORMATIVE TIME-LAG STUDIES WITH HIGH-FREQUENCY DISCHARGES

By A. W. BRIGHT, Ph.D., B.Sc.(Eng.), Graduate, and H. C. HUANG, Ph.D., B.Sc.(Eng.).

(*The paper was first received* 20*th March and in revised form* 27*th May*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in August*, 1954.)

SUMMARY

Experiments with discharges between non-uniform electrodes at frequencies in the region of 1 to 15 Mc/s are described. It was found that, above a certain critical frequency, the discharge was able to cross the gap completely, whereas the discharge at direct current or 50 c/s with the same electrodes would take the form of a localized corona envelope.

By using pulse techniques, it was possible to measure the formative time-lags for h.f. discharges, using concentric-cylinder and parallel-wire electrodes. A lowering of the breakdown voltage of some 13% was observed to take place above the critical frequency.

A tentative qualitative explanation is proposed of some of the observed results.

(1) INTRODUCTION

Electrical discharges between electrodes having a non-uniform field distribution display interesting and unusual properties when the frequency of the applied voltage is in the region of 1 to 15 Mc/s. One of the reasons for this is that positive ions produced in the discharge, owing to their relatively lower mobility compared with electrons, have a transit time which may be less than the period of a half-cycle of applied voltage. Such ions are not swept to the electrodes but can remain in the gap, and indeed their number increases with each succeeding half-cycle, resulting in the accumulation of an appreciable space-charge.[1,2,3]

One property of such an r.f. discharge is that complete breakdown of the gap can take place at a frequency of a few megacycles per second whereas at 50 c/s only a stable corona discharge would be observed with the same electrode geometry. There is clearly a region in the frequency band where the discharge changes from the corona region to a region where a discharge completely crosses the gap between electrodes. In the work which is described below, an investigation was made of this region using concentric-cylinder and parallel-wire electrode configurations.

The work formed part of a more general study of the r.f. discharge in gases at Queen Mary College, University of London.

(2) APPARATUS

The apparatus consisted basically of an r.f. oscillator which was coupled inductively to a tuned circuit, the test electrodes contributing part of the tuning capacitance. With this arrangement it was possible to produce a voltage of up to 8 kV at frequencies from 1 to 15 Mc/s. The voltage across the electrodes was measured by a simple diode-voltmeter and an electrostatic voltmeter. In order that pulse voltages would be accurately measured, considerable care was taken with insulation of the voltmeter. The discharge time-constant of the voltmeter was about 20 min. Using the criteria developed by Burges,[4] it was arranged that the voltmeter error when measuring 50 microsec pulses at 50 c/s repetition frequency would be extremely small. The voltmeter was carefully calibrated at 50 c/s by three inde pendent methods, and the overall error was estimated to be les than $\pm 2\%$.

The applied r.f. voltage was either in pulses with a repetitio frequency of 50 c/s or alternatively in non-recurrent pulses o pulses with a repetition rate of one per second. The duratio of these pulses could be varied from about 10 to 2 000 microsec

Timing and triggering circuits were incorporated so that th envelope of the r.f. voltage across the test electrodes could b displayed on one beam of a double-beam oscillograph, a trai of timing pulses being displayed on the second beam. Th interval between the sharp timing pulses could be either 20 o 50 microsec.

(3) ELECTRODE CONFIGURATION

The concentric-cylinder electrodes consisted of a tinned copper wire suspended vertically inside a copper cylinder, th tube being earthed and the voltage applied to the wire. Th tube was 46 cm long and 5 cm in diameter. The ends of th tube were flared to avoid sharp edges which might lead t breakdown. The electrodes were mounted inside a light-tigh box, and discharges could be observed by means of an incline mirror arranged some distance vertically below the wire. Fres air could be drawn slowly through the box by means of a smal electric fan. This removed any accumulation of unwanted gase such as ozone, which were present after periods of corona.

In order to facilitate observation and photography, a secon electrode system was constructed with a short cylinder of 3·8 cm diameter and 5 cm long. In this case, the wire was coated wit polystyrene over all except for a short section at the mid-poin of the wire. In this way the discharge was localized and coul be placed at the focal point of a camera lens.

For tests with parallel wires a third arrangement was used One of the wires was fixed on a block of polystyrene which wa so constructed that the wire could be tightened with screw The second wire was also fixed on a polystyrene block. Th two blocks were suitably supported so that the wires could b made exactly parallel and the distance between the wires coul be adjusted and measured by means of a micrometer head.

For wire spacings from 0·5 to 10 mm, the greatest divergen from the parallel was less than 0·1%.

The wires used in the concentric-cylinder tests were tinne copper fuse wires. The greatest divergence from the nominal di meter of these wires was, with one exception, $\pm 1{\cdot}5\%$. Micr scopic examination showed that the wire surface was extreme smooth with few irregularities.

In the parallel-wire tests, both tungsten and tinned-copper wir were used. Tungsten wire, after being annealed and polishe was convenient to use because it could withstand a number discharges without melting.

The usual methods were used to remove grease and dirt fro electrode surfaces. Before any measurements were made, t wires were "conditioned," by allowing h.f. corona to take pla over the wire surface for some minutes. This removed any sm irregularities which would lead to a local field intensification.

Correspondence on monographs is invited for consideration with a view to publication.
Dr. Bright is at the Imperial College of Science and Technology, University of London.
Dr. Huang was formerly at Queen Mary College, University of London.

(4) TIME-LAG MEASUREMENTS

(4.1) Concentric Cylinder

First tests were made with concentric-cylinder electrodes. At direct current and at 50 c/s with the wire diameters used (0·234mm–0·457mm), breakdown took the form of a very faint corona discharge on the wire surface. The onset of corona could only be detected with accuracy by connecting a 100000 ohm resistor in series with the outer cylinder and amplifying the voltage across this resistor. When the applied voltage was increased, the diameter and intensity of the corona envelope increased.

When pulses of 10 Mc/s at 50 pulses a second were applied to the electrodes the breakdown phenomenon was quite different. As soon as the critical voltage was reached, a bright filamentary discharge bridged the gap and the wire was covered with brilliant blue-violet streamers, most of which bridged the gap between wire and cylinder. The onset voltage was well defined.

To reduce the statistical lag, or in other words, to provide a source of free electrons in the gap, a 0·8 millicurie capsule of cobalt 60 was placed near the electrodes.

A sketch of a typical oscillogram corresponding to this discharge is shown in Fig. 1. Point A is the beginning of the

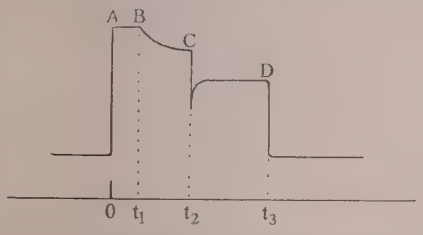


Fig. 1.—A sketch of a typical breakdown pulse.

pulse, point B corresponds to the onset of a discharge and time t_1 is the statistical lag. With no irradiation, this time varied in a random manner and sometimes no discharge occurred. With adequate irradiation this time was too small to measure with the existing apparatus and was certainly less than 1 microsec.

At point C, breakdown was complete at time t_2. The interval $(t_2–t_1)$ was taken to be the formative lag. It was found that this time had a definite minimum value, which could be determined either by applying about one hundred pulses and observing the minimum value of t_2 (t_1 was normally zero) or by applying 50 pulses a second and noting t_2 on the oscillogram of the superimposed discharges. The discrepancy between the results obtained from the two methods was about 10%. As was to be expected, the time-lag decreased rapidly with increased over-voltage, but, if the breakdown voltage was determined accurately by raising the applied voltage slowly, the time-lag results could be reproduced readily.

The results, shown in the curves of Fig. 2, indicate that, at frequencies above about 4 Mc/s, the time-lag tended to a constant value which was determined by the wire diameter.

The validity of the procedure adopted was proved by applying a pulse whose amplitude was equal to the measured breakdown voltage, but whose duration was less than the measured formative time-lag. In this case, the discharge filament did not completely bridge the gap and it was possible to photograph a mid-gap "chopped" spark.[5] By photographing streamers in this way, using pulses at a repetition frequency of 50 c/s, varying the pulse duration and measuring the maximum length of the discharge, it was possible to estimate the speed of progress of the discharge.

As the frequency was reduced below 4 Mc/s, the time-lag began to rise quite rapidly and tended to very large values below

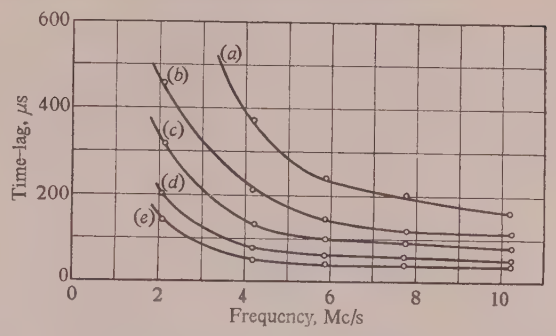


Fig. 2.—Minimum time-lag between application of r.f. pulses and complete breakdown, at various frequencies.

Concentric-cyclinder electrodes: radius of outer cylinder, 25mm.
(*a*) Inner-wire diameter, 0·234mm.
(*b*) 0·274mm. (*c*) 0·316mm. (*d*) 0·376mm. (*e*) 0·457mm.

3 Mc/s. In fact there was a critical frequency below which the discharge was unable to bridge the gap, and so the time-lag was infinite.

The variation of breakdown voltage with frequency is shown in Fig. 3. It will be noted that there is a well-defined region over which the breakdown voltage decreased with increasing

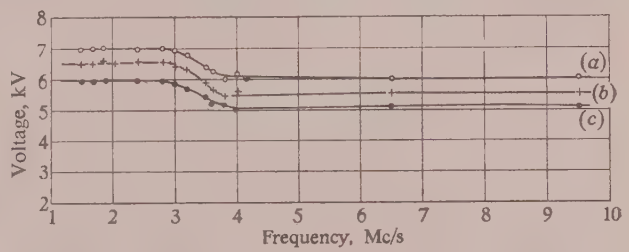


Fig. 3.—Variation of r.f. breakdown voltage with frequency for concentric-cylinder electrodes.

Radius of outer cylinder, 25mm.
(*a*) Inner-wire diameter, 0·316mm.
(*b*) 0·274mm. (*c*) 0·234mm.

frequency. This critical frequency range is close to that at which the time-lag was observed to change rapidly with frequency. As will be gathered from the discussion, there is good reason to suppose that the two phenomena are closely related.

With a wire of 0·316mm diameter, a cylinder of 38mm diameter and frequency of 3·4 Mc/s, the speed of progress of the discharge across the gap was about 4×10^4 cm/sec.

(4.2) Parallel Wires

With parallel wires an attempt was made to specify the breakdown voltage more accurately. Pulses with a repetition rate of one per second were applied to the gap, and the breakdown voltage was taken to be the value corresponding to the gap breaking down for 50% of the applied pulses. The time-lag was then determined in a manner similar to that for the previous case.

Typical results of these tests are shown in Fig. 4. It will be noted that the curves are of the same general shape as those in Fig. 2, but the minimum value of time-lag was much less with twin wires than with concentric cylinders.

Once more it was found that, provided the frequency was low enough, the discharge was unable to bridge the gap.

If the mean speed of progress of the streamer was calculated by dividing the gap distance by the time-lag, a similar result was obtained as in the case of concentric cylinders.

For a wire diameter of 0·10mm and a wire spacing of 5mm,

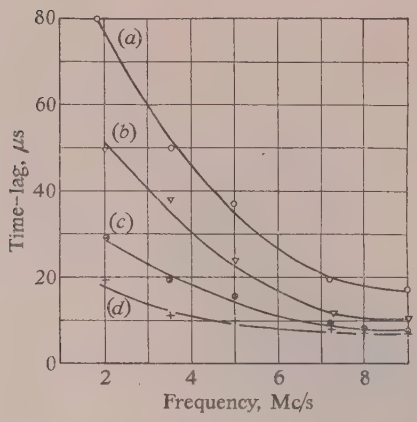


Fig. 4.—Formative time-lag of parallel-wire electrodes.
Wire diameter = 0·20 mm.
Spacing between wires:
(*a*) 8 mm. (*c*) 6 mm.
(*b*) 7 mm. (*d*) 5 mm.

the streamer speed at 3·5 Mc/s was $7{\cdot}8 \times 10^4$cm/sec, which is the same order as the value obtained with the concentric-cylinder arrangement. It was found that the time-lag could be shortened considerably if large over-voltages were used. This is a well-known phenomenon in d.c. time-lag studies.

(5) DISCUSSION OF RESULTS

The main interest of these results lies in the marked difference between the breakdown phenomena of a given gap under r.f. and d.c. conditions. The electrode geometries were such that at d.c. or 50 c/s breakdown would take the form of a gentle corona discharge at the surface of one or both of the wires. With the r.f. discharge, however, complete breakdown would usually occur, provided that the frequency were high enough, since the discharge had to progress into a region of progressively lower electric field. The explanation of the observed results must almost certainly be connected with the accumulation of positive ions which was mentioned in the introduction; but any attempt at a detailed analysis would encounter very great difficulty.

Before discussing the possible mechanism by which the h.f. discharge propagates, it is convenient to examine the mechanism of positive and negative d.c. corona.

With negative-wire corona the progress of electron avalanches from the wire is eventually inhibited by positive space-charge which is established as the avalanche moves into the region of decreasing electric field. Under certain conditions, the space charge can interrupt the discharge, and relaxation oscillation of the type observed by Trichel can occur.

With positive-wire corona the mechanism is quite different. Now, the positive space-charge tends to extend the region of high electric field into the gap, as the space-charge field and applied field are additive. The discharge can move out across the gap until the space-charge field can no longer outweigh the reduction in field determined by the electrode geometry. A stable visible corona discharge is then established around the wire.

At low frequencies, positive and negative corona will take place alternately at the successive peaks of applied voltage. Again the corona is localized about the surface of the wire. During the intervals between voltage peaks, most of the positive ions and free electrons remaining in the gap are swept to the electrodes.

Having examined the reasons for the localized nature of l.f. and d.c. corona, we may now attempt to explain why the h.f. corona can completely cross the gap between electrodes.

It is fairly well established that positive ions are able to perfor mid-gap oscillations in uniform field gaps under certain condi tions of electric field, gap distance and frequency. The fact tha these ions tend to accumulate in the gap to form a space charg is considered to be the mechanism which causes the well-know "critical-gap"[1,2,6] phenomenon and the observed lowering c breakdown voltage which occurs at frequencies in the region c 1 to 15 Mc/s with gap spacings of 0·5–5 mm.

A similar critical-gap phenomenon occurs with concentri cylinder electrodes. This is shown clearly in the curves c Fig. 4. One of the authors[2] has developed a semi-empirica equation for this case giving the frequency at which positiv ions can perform oscillations close to the central wire withou touching it. This is of the form:

$$f_c = \frac{AV}{(B^2 - r^2)\log(R/r)}$$

where V is the peak value of the applied voltage, A is a cor stant, B is a factor which depends on the gap geometry and th gas, and R, r are the radii of the outer and inner cylinde respectively.

Above this critical frequency, positive-ion space charge wi accumulate in the gap and there is a lowering of the breakdow voltage. With the present dimensions the critical frequency wa between 2 and 3 Mc/s with concentric cylinders.

A possible mechanism envisaged by the authors is the fo lowing. For simplicity, we will consider the case of a thin wir surrounded by a concentric cylinder.

Fatechand[3] showed that, if the frequency were high enough positive ions could be detected in the gap before the voltage wa high enough to cause a visible discharge. The resulting positiv space-charge would enhance the field at the wire surface at th peak value of the negative half-cycle. Thus the value of applie voltage necessary to cause onset of a discharge would be lowe than that which would be required in the absence of space charg The magnitude of this effect is seen in Fig. 3 to be a lowerin by about 13% of the l.f. corona voltage value.

During the interval between the peak of the negative half-cycl and that of the positive half-cycle (about 10^{-7}sec) not all th positive ions would be swept to the electrodes. The ions whic remained would find themselves fairly close to the surface o the wire when it reached its maximum positive voltage.

At the positive voltage peak, a discharge would again proce from the wire. At low frequencies, we saw that the distanc which the corona could propagate was largely determined b the magnitude of the positive space-charge produced. At hig frequencies, it is seen that the positive ions produced by th electron avalanche would be augmented by the positive ions le over from the preceding negative half-cycle. Further positiv ions would be produced during the next negative half-cycle, bu the discharge would not progress further across the gap becaus of the inhibiting effect of the space charge. However, when th wire was again positive, the discharge would again move furthe across the gap, since the applied and space-charge fields would add

In this way, the discharge channel would move from wire t outer cylinder in a series of steps, making up for the progressivel lower field due to electrode geometry by its increasing positiv space-charge. Consideration of the results show that, at 10 Mc/ some 2 000 half-cycles were required for complete breakdow Photographs showed that the progress of the discharge was ver uncertain and forked, which is a slight indirect evidence of space-charge effect.

No mention has been made of the influence of electrons, apa from the electron avalanches.

The frequencies under consideration were too low to allow th

ctron transit time to be of much importance. Any electrons hich survived a half-cycle would tend to neutralize the positive ace-charge.

Some indirect evidence in favour of the proposed mechanism to be found in the very short time-lags found with h.f. dis-arges[7] in uniform fields. Here, the discharge is not required move into a region of steadily decreasing field and no rela-ely slow step-by-step progress would be expected.

It is not suggested that the above explanation is anything ore than a reasonable possibility. In this field of gas dis-arges, it is difficult to do more than offer tentative explanations r the observed phenomena, as quantitative analysis of the oblem is at the moment virtually impossible owing to the many known parameters.

The tentative explanation proposed by the authors can only approximately true, since the effect of electron space-charges s not been considered, but it is thought that the influence of ese would be small.

It might be possible to go a little further towards solving the oblem by study of the negative d.c. corona. With negative rona it is known that Trichel pulses of current of short duration cur, owing to the influence of positive space-charges. From ese negative corona studies it might be possible to measure e magnitude of space charges involved. This information ould be valuable in any more accurate explanation of the h.f. scharge.

(6) ACKNOWLEDGMENTS

The work was carried out in the Electrical Engineering department of Queen Mary College. The authors wish to thank the British Electrical and Allied Industries Research Association which has been a constant source of help in many ways. They are particularly grateful to Dr. W. A. Prowse of Durham University for his advice and interest in the problem.

(7) REFERENCES

(1) Böcker, H.: "Die Durchschlagsenkung bei Hochfrequenz," *Archiv für Elektrotechnik*, 1937, **31**, p. 166.
(2) Bright, A. W.: "Corona and Breakdown at Frequencies up to 12 Mc/s," E.R.A. Report Ref. L/T229.
(3) Fatechand, R. R. T.: "Electrical Breakdown of Dielectrics at Radio Frequencies," Ph.D. Thesis, University of London, 1951.
(4) Burges, R. E.: "The Response of a Linear Diode-Voltmeter to Single and Recurrent R.F. Impulses of Various Shapes," *Journal I.E.E.*, Part III, 1948, **95**, p. 106.
(5) Bright, *ibid.*
(6) Lassen, H.: "Frequenzabhaegigkeit der Funkenspannung in Luft," *Archiv für Elektrotechnik*, 1931, **25**, p. 322.
(7) Huang, H. C.: "Radio-Frequency Discharges between Parallel-Wire and between Parallel-Plane Electrodes," Ph.D. Thesis, University of London, 1953.

621.391 : 621.38

Monograph No. 106
Sept. 19

THE RESPONSE OF A NON-LINEAR SYSTEM TO RANDOM NOISE

By W. E. THOMSON, M.A.

(*The paper was first received* 12*th April, and in revised form* 30*th June*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in September*, 1954.)

SUMMARY

A formula for the auto-correlation function of the output of a non-linear system, the input being random Gaussian noise, has been given in the literature for the case where the non-linear output/input relation is expressed as a power series. This formula can be simplified; the new version emphasizes the fact that intermodulation products of different orders are uncorrelated, but that contributions to the intermodulation product of a given order from different terms of the power series are completely correlated.

If the output is expressed as a series of Hermite polynomial functions of the input, rather than as a power series, each term of the series gives rise to intermodulation products of one order only.

LIST OF SYMBOLS

V_0 = R.M.S. value of input noise.
$x(t)$ = Random Gaussian noise of unit r.m.s. value.
$y(t)$ = Output of non-linear system.
$\psi(\tau)$ = Auto-correlation function of $x(t)$.
$\Psi(\tau)$ = Auto-correlation function of $y(t)$.
μ_{mn} = Bivariate moment of order m, n for a bivariate normal distribution.
$H_n(x)$ = Hermite polynomial of order n. There are variations in the definition; that used in the paper is $H_n(x) = (-1)^n \varepsilon^{x^2/2} \dfrac{d^n}{dx^n}(\varepsilon^{-x^2/2})$.
$[f(t,\tau)]_{av}$ = Average value of $f(t,\tau)$ over t for a fixed value of τ.

(1) INTRODUCTION

The response of a non-linear system to a random noise input is of sufficient interest to have been dealt with in many papers.[1,2,5,6] The intention of the present paper is, first, to point out a simplification of a previously given formula for the case when the output is expressed as a weighted sum of powers of the input, and secondly, to demonstrate that the expression of the non-linear output/input relation as a weighted sum of Hermite polynomials leads more directly to the division of the output into intermodulation products of various orders.

The mathematical analysis of the paper has been merely outlined. In Section 3.1 in particular, a plausible demonstration, linking up with the more familiar power-series expansion, has been used, rather than a more formal proof *ab initio*.

The convention regarding auto-correlation functions differs slightly from that of most other authors. The input noise is taken as $V_0 x(t)$, where V_0 is the r.m.s. value of the noise and hence covers the magnitude and dimensions of the noise; $x(t)$ is a dimensionless noise function of unit r.m.s. value and its auto-correlation function is therefore normalized, i.e. $\psi(0)$ equals unity.

(2) OUTPUT AS POWER SERIES

Let the input $V_0 x(t)$ be random Gaussian noise of r.m.s. value V_0 and the output be

$$y(t) = \sum_{n=0}^{\infty} a_n [V_0 x(t)]^n \quad . \quad . \quad . \quad . \quad . \quad (1)$$

We determine $\Psi(\tau)$, the auto-correlation function of $y(t)$; th power spectrum of $y(t)$ can then be obtained from $\Psi(\tau)$ by th usual methods:

$$\Psi(\tau) = [y(t)y(t+\tau)]_{av}$$
$$= \sum_{m,n=0}^{\infty} a_m a_n V_0^{m+n} \{[x(t)]^m [x(t+\tau)]^n\}_{av} \quad . \quad (2$$

where $\{[x(t)]^m [x(t+\tau)]^n\}_{av}$, a time-average for fixed τ, repre sents an auto-correlation function for $m = n$, and a cross correlation function for $m \neq n$; each such cross-correlatio appears twice effectively since the same average is obtained b interchanging m and n.

Since the noise is Gaussian

$$\{[x(t)]^m [x(t+\tau)]^n\}_{av} = \mu_{mn}$$

the bivariate moment of order m, n for a bivariate normal (i.e Gaussian) distribution with unit variances and correlatio coefficient $\psi(\tau)$, the auto-correlation function of $x(t)$. It can b shown[3] that

$$\mu_{mn} = \mu_{nm}$$
$$= 0 \qquad m + n \text{ odd}$$
$$= \sum_{j=0}^{\leqslant \frac{1}{2}m} \frac{m!n![\psi(\tau)]^{m-2j}}{2^{\frac{1}{2}(n-m)+2j} j!(m-2j)![\frac{1}{2}(n-m)+j]!} \qquad m + n \text{ even}, \quad n \geqslant m$$
$$. \quad . \quad . \quad . \quad . \quad (3$$

Eqns. (2) and (3) represent the previously given result and ar equivalent to Middleton's[1] eqns. (8.4) and (8.5) and to a equation of Section 2.2 of Clavier's paper,[2] the latter collectin terms in powers of $\psi(\tau)$ as far as the fourth.

The term in $[\psi(\tau)]^n$ has a physical meaning. As shown, fo example, by Rice,[6] random Gaussian noise can be considered a the infinite sum of elementary sinusoidal components, and nt order intermodulation products of these components in th output of a non-linear system turn out to give rise to the ter in $[\psi(\tau)]^n$ in the auto-correlation function of the outpu Let us now determine the term in $[\psi(\tau)]^r$ from eqns. (2) and (3) r will be used for the moment to avoid confusion with the n in m, Contributions are obtained with $m = r + 2j$, $n = r + 2k$, fo all non-negative integral values of j and k; hence

$$\Psi(\tau) = \sum_{r=0}^{\infty} [\psi(\tau)]^r \sum_{j,k=0}^{\infty} \frac{(r+2j)!(r+2k)!}{2^{j+k} j!r!k!} a_{r+2j} a_{r+2k} V_0^{r+2j+r+}$$

It can now be seen that the inner summation, with a factor removed, is in the form of a perfect square, and we get the ne formula, in which r can now be replaced by n

$$\Psi(\tau) = \sum_{n=0}^{\infty} n![\psi(\tau)]^n \left[\sum_{j=0}^{\infty} \frac{(n+2j)!}{2^j j!n!} a_{n+2j} V_0^{n+2j} \right]^2 . \quad ($$

Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Thomson is at the Post Office Research Station.

t may be helpful to give the explicit form for powers up to the ifth; it is

$$\begin{aligned}\Psi(\tau) = {} & [a_0 + a_2V_0^2 + 3a_4V_0^4 + \ldots]^2 \\ & + \psi(\tau)[a_1V_0 + 3a_3V_0^3 + 15a_5V_0^5 + \ldots]^2 \\ & + 2[\psi(\tau)]^2[a_2V_0^2 + 6a_4V_0^4 + \ldots]^2 \\ & + 6[\psi(\tau)]^3[a_3V_0^3 + 10a_5V_0^5 + \ldots]^2 \\ & + 24[\psi(\tau)]^4[a_4V_0^4 + \ldots]^2 \\ & + 120[\psi(\tau)]^5[a_5V_0^5 + \ldots]^2 \quad . \quad . \quad . \quad . \quad . \quad (5)\end{aligned}$$

n which the first line gives the d.c. component of the output, he second the undistorted output, and third and subsequent ines, second- and higher-order intermodulation products.

This simplified form of the expression for $\Psi(\tau)$ is associated vith the physical fact that, for a Gaussian input, the contribuions to intermodulation products of a given order from the arious powers concerned are completely correlated and add oltagewise, as has been pointed out by Lewin[4] (who refers to "coherent low-order groups"); intermodulation products of lifferent orders are uncorrelated and add powerwise.

(3) OUTPUT AS SERIES OF HERMITE POLYNOMIALS

(3.1) General

The relation between output and input in a non-linear system nay be expressed in a perfectly general way as

$$y(t) = f[V_0x(t)]$$

vhere f is some function, possibly in analytic form, possibly an xperimental curve. The power series of eqn. (1) is often used s a simple approximation to some given curve, but other xpansions are possible. Eqn. (4) shows that the power series s not a very good expansion if we want to deal with intermodulation products of various orders in the output, since all powers up to infinity occur in any one term. It may be asked vhether it is possible to express y as a sum of functions, each f which is associated with intermodulation products of a given rder? The answer is yes. We can find functions such that if

$$y(t) = c_0H_0[x(t)] + c_1H_1[x(t)] + c_2H_2[x(t)] + \ldots \quad . \quad (6)$$

hen $$\Psi(\tau) = c_0^2 + c_1^2\psi(\tau) + c_2^2 2![\psi(\tau)]^2 + \ldots \quad . \quad . \quad (7)$$

A comparison of eqns. (7) and (4) shows that

$$c_n = \sum_{j=0}^{\infty} \frac{(n+2j)!}{2^j j! n!} a_{n+2j} V_0^{n+2j} \quad . \quad . \quad . \quad (8)$$

The functions $H_n(x)$ have now to be found to make eqns. (1) nd (6) equivalent; grouping the right-hand side of eqn. (6) to orrespond to powers of x in eqn. (1) gives

$$\begin{aligned} H_0(x) &= 1 \\ H_1(x) &= x \\ H_2(x) + H_0(x) &= x^2 \\ H_3(x) + 3H_1(x) &= x_3 \\ &\vdots \\ \sum_{j=0}^{\leqslant \frac{1}{2}n} \frac{n!}{2^j j!(n-2j)!} H_{n-2j}(x) &= x^n \quad . \quad . \quad . \quad (9) \end{aligned}$$

which may be solved to give

$$\begin{aligned} H_0(x) &= 1 \\ H_1(x) &= x \\ H_2(x) &= x^2 - 1 \\ H_3(x) &= x^3 - 3x \\ &\vdots \\ H_n(x) &= \sum_{j=0}^{\leqslant \frac{1}{2}n} \frac{(-1)^n n!}{2^j j!(n-2j)!} x^{n-2j} \quad . \quad . \quad . \quad (10) \end{aligned}$$

which can be recognized as the Hermite polynomials.[3] So, by a well-known property of these polynomials,

$$c_n = \frac{1}{n!(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} f(V_0x)\varepsilon^{-x^2/2} H_n(x)dx \quad . \quad . \quad (11)$$

which allows c_n to be determined direct from the given non-linear relation, by numerical integration if need be. The total power in nth-order intermodulation products (the d.c. and undistorted components being orders 0 and 1 respectively) is $n!c_n^2$.

The expansion in Hermite polynomials can sometimes be possible when the ordinary Taylor power series is not, because of discontinuities. The example in Section 3.2 illustrates this.

(3.2) Noise through a Half-Wave Rectifier

As an example of the method of Section 3.1, consider a half-wave rectifier for which

$$\begin{aligned} f(x) &= 0 \quad x \leqslant 0 \\ &= x \quad x \geqslant 0 \quad . \quad . \quad . \quad . \quad . \quad (12) \end{aligned}$$

so that

$$c_n = \frac{1}{n!(2\pi)^{\frac{1}{2}}} \int_0^{\infty} V_0 x\epsilon^{-x^2/2} H_n(x)dx \quad . \quad . \quad . \quad (13)$$

the fact that

$$\varepsilon^{-x^2/2} H_n(x) = (-1)^n \frac{d^n}{dx^n}(\varepsilon^{-x^2/2})$$

allows the integral to be evaluated by integration by parts twice with special treatment for c_0 and c_1 and yields

$$\begin{aligned} c_0 &= V_0/(2\pi)^{\frac{1}{2}} \qquad c_1 = \tfrac{1}{2}V_0 \\ c_n &= \frac{(-1)^{m-1}V_0}{2^m m!(2m-1)(2\pi)^{\frac{1}{2}}} \quad \begin{matrix} n \text{ even}, \geqslant 2, \\ \text{and } = 2m \end{matrix} \quad c_n = 0 \quad \begin{matrix} n \text{ odd}, \geqslant 3 \end{matrix} \\ & \quad . \quad . \quad . \quad . \quad (14) \end{aligned}$$

The resulting series for $\Psi(\tau)$ is a special case of a series of Middleton's,[5] and may also be recognized as the expansion of the form given by Rice:[6]

$$\Psi(\tau) = \frac{V_0^2}{2\pi}\left[\left\{1 - [\psi(\tau)]^2\right\}^{\frac{1}{2}} + \psi(\tau) \arccos\left[-\psi(\tau)\right]\right] \quad (15)$$

(4) ACKNOWLEDGMENTS

The author is indebted to Mr. E. C. H. Seaman, who pointed out the possibility of the type of expansion dealt with in Section 3.1, and to the Engineer-in-Chief of the Post Office for permission to make use of the information given in this paper.

(5) REFERENCES

(1) MIDDLETON, D.: "Some General Results in the Theory of Noise through Non-linear Devices," *Quarterly of Applied Mathematics*, 1948, **5**, p. 445.

(2) CLAVIER, P. J.-M.: "Distortion non-linéaire d'un signal aléatoire; application à la modulation en fréquence," *Câbles et Transmission*, 1953, **7**, p. 293.

(3) KENDALL, M. G.: "The Advanced Theory of Statistics" (Charles Griffin and Co., Ltd., 1945), Vol. 1.

(4) LEWIN, L.: "Interference in Multi-Channel Circuits," *Wireless Engineer*, 1950, **27**, p. 294.

(5) MIDDLETON, D.: "The Response of Biased Saturated Linear and Quadratic Rectifiers to Random Noise," *Journal of Applied Physics*, 1946, **17**, p. 778.

(6) RICE, S. O.: "Mathematical Analysis of Random Noise —Part II," *Bell System Technical Journal*, 1945, **2** p. 46.

21.374.32 : 681.142

Monograph No. 107 M
Sept. 1954

THE FUNCTION OF BASIC ELEMENTS IN DIGITAL SYSTEMS

By C. B. SPEEDY, Ph.D., B.E.

(*The paper was first received* 11*th March, and in revised form* 1*st June*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in September*, 1954.)

SUMMARY

It is shown that electronic digital calculating machinery, and stems which represent information in digital form, may in principle constructed from an assemblage of three basic elements. These e a bistable element for the storage of a digit, a gate for controlling e flow of digits, and a diode for controlling the direction of flow of gits. The properties and use of an element which consists of a stable element and two gates are described, and it is then shown that e functions of this element may be realized in the form of a practical cuum tube based upon beam-deflection principles. This tube does t lend itself to the construction of a practical large-scale store, but ther to extensive use in the control and arithmetic sections of a gital machine.

(1) INTRODUCTION

The principles underlying the operation of automatic digital culating machinery and devices which handle information in gital form were firmly established 100 years ago when Charles bbage detailed the design for a machine to perform automatic mputation. At that time manufacturing techniques were dequate, and as a consequence the practical realization of machine was delayed until the mass-production methods of present century had been established.

The past decade has seen the successful construction and eration of both the electro-mechanical and electronic forms automatic digital calculating machinery. The electronic chines, which are significant because of their very high speed operation, have been made possible largely by the availability conventional triodes (and pentodes, etc.) and the techniques veloped primarily for use in the communication industry. e use of the triode in this manner depends not upon its linear plifying characteristics but rather upon the non-linearities sent at the extremities of its normal operating range. This ggests that the conventional triode may not necessarily be best tube for this application, and that it may be possible to velop a tube which is more directly suited to the storage and itching requirements of high-speed digital machinery.

The idea of building special tubes for use in electronic digital culating machinery is not new. In this field significant advances ve been made in the development of numeric storage tubes. th Haeff[1] and Rajchman[2] have constructed tubes for storing ny hundreds of binary digits by means of electric charges ained in small cells upon an insulating surface. The demonation by Williams[3] of the use of a conventional cathode-ray be for digital-storage purposes has fostered the development tubes free from screen imperfections. Pulse counting in nputers has led to the development of tubes for performing s function. The Hollway[4] decimal counter tube, which pends for its operation upon the location of an electron beam one or other of ten stable positions, is an example. Another proach has led to the development of special tubes for perming certain complicated computing functions. In the Computron[5] an attempt has been made to devise a tube with several hundred electron beams, for computing continuously the product of two binary numbers. Dr. Katz at Toronto has developed a group of special tubes which are of simple and conventional construction. He has shown that by adopting a particular form of construction, a variety of characteristics can be obtained by slight modifications of the electrode shapes. In 1946 Sharpless[6] suggested a course of development based on the single-beam bistable tube, observing that the use of beam deflection rather than grid control of the target current would eliminate the need for potential-dividing circuits when connecting the output of one tube to the input of another. Jonker[7,8] has more recently outlined the electrical and mechanical requirements of switching tubes for use in telephony and possibly in calculating machines. He has emphasized the value of ribbon-like electron beams in this application, and also the use of a simple and conventional form of construction. As opposed to previous efforts the present work has aimed at reducing the quantity and variety of components in digital machinery by endeavouring to devise a general-purpose tube to perform the basic storage and switching functions in such machinery.

The functions of the simplest basic elements may be determined by considering the necessary and sufficient combinations of two binary numbers in a digital system. These combinations are the logical ones of conjunction and disjunction, the former corresponding to digit-by-digit multiplication and the latter to logical summation. From basic elements performing these functions, together with a third element for storing a binary digit, devices can be constructed in a simple manner for performing the operations of addition, multiplication, counting, etc. Using the terminology of digital machinery, these elements are termed a "gate" for performing the process of conjunction, a pair of "diodes" for disjunction, and a "bistable" element for the storage of a digit.

In principle the bistable element may be one or other of two kinds: either a static device in which the value of the digit is determined by the direct potential of its output, or a dynamic device whose state is indicated by either the presence or absence of a continuous succession of output pulses. Since the output signals from these elements are always applied to gate or diode elements where the logical operations are performed, it is common to use the static device in preference to the dynamic one because of the practical difficulty which is found in matching the output pulses (in both voltage and time) from two bistable elements of dynamic form.

If these elements are to be realized in practical form it is only necessary to consider the bistable and gate elements, since the diode is currently available in simple form. The observation that the output of a bistable element of static form is always connected to the input of a gate suggests a composite element consisting of a bistable element permanently connected to a gate. Symmetry suggests, as a further step, the use of two gates permanently connected to the opposite outputs of a bistable element such that one gate is open to the flow of information and the other closed when the digit stored is unity, and vice versa

orrespondence on Monographs is invited for consideration with a view to lication.

r. Speedy is in the Electrical Engineering Dept., Imperial College of Science and nology.

when it is zero. Although the symmetrical arrangement is not essential for the composite element, there are several devices, such as shift registers, counters, encoders, etc., where two gates are required for each bistable element. In these cases the inclusion of the second gate in the composite element does add to the element's complexity. As opposed to this, the total number of components in the device is reduced, and as seen later, the addition of the second gate does not appreciably complicate the construction of the vacuum tube designed to perform these functions. An element consisting of a bistable element and two gates is referred to here as a "binary gating" element.

To perform the function of the binary-gating element a binary-gating tube has been built. This depends for its operation upon the independent deflection of three ribbon-like electron beams and the collection of their current in a target assembly. Deflection control of the target current rather than grid control, as used in triodes, permits the output target of one tube to be connected directly to the input deflector of another without the use of potential-dividing circuits. The use of three ribbon-like electron beams rather than a single circular beam leads to a desirable electrode arrangement consisting of three simple cells each of similar construction. The beams are produced in a cylindrical form of the Pierce-type electron gun, which has the advantages of simplicity and relative ease of design. In the output part of the tube the possibility of using secondary emitting targets which are electrically floating has been considered. The main difficulty encountered was that of obtaining a high secondary-emission ratio over a long period of time. Problems also arise in the use of such tubes in supplying the secondary-emission current when connecting the output of one tube to the input of another. Consequently the use of secondary emission was abandoned, and the beam currents were used directly by suppressing the secondary emission at the targets in the normal manner.

(2) THE BINARY-GATING ELEMENT AND ITS APPLICATIONS

The application of the binary-gating element and the diodes to the construction of a digital system is indicated by considering in detail their use in a typical portion of an arithmetic unit of an electronic digital calculating machine. The example chosen is that of a shift register with an associated binary counter, as is used in serial binary machines for the temporary storage of numbers. Such a device has not been constructed from binary-gating tubes; it is considered here simply to indicate the manner in which the tubes could be used in a practical system.

The symbol used for representing the binary-gating element is shown in Fig. 1. Terminal 1 is connected to the input of the

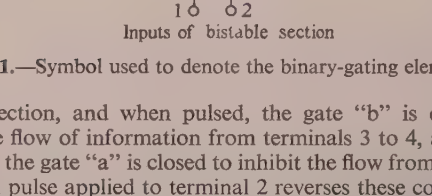


Fig. 1.—Symbol used to denote the binary-gating element.

bistable section, and when pulsed, the gate "b" is opened to permit the flow of information from terminals 3 to 4, and at the same time the gate "a" is closed to inhibit the flow from terminals 5 to 6. A pulse applied to terminal 2 reverses these conditions.

Although the symbol has been considered in terms of pulse representation of digits, the binary-gating tube is so designed that direct connections can be made between tubes. Because of this property, the use of the tube may be extended to the application where the binary-digit representation is by means two different voltage levels, rather than by pulses.

Continuously-generated timing pulses, termed "clock puls p_c, are separated by intervals of time, termed "clock periods" duration T. By the use of extra timing pulses having the sa repetition frequency as the clock pulses but separated from th by some fraction of a clock period, binary-gating elements n be used as delay elements. These additional pulses are refer to here as "quarter pulses" p_a, "shift pulses" p_s, and "thr quarter pulses" p_b, all of which are shown in Fig. 2. T occur at one-quarter, one-half and three-quarters of a cl period, respectively, after the clock pulses.

Digits are transmitted by means of two lines, known as "D-bus" and "1/D-bus." Digit 1 is represented by a pulse the D-bus and no pulse on the 1/D-bus, digit 0 being represen by the reverse conditions. The double bus is not essential, is used in this case because it tends to simplify the circuits.

For an example of the use of the binary-gating eleme consider a shift register with a capacity for 40 binary dig together with an associated counter for generating the pul to control the flow of digit pulses to and from the register. indicated in Fig. 2, a number entering the register upon input bus (consisting of both the D-bus and the 1/D-bus) preceded by a "write pulse" upon the "write-pulse bus" by interval of one clock period. An outgoing number upon same digit bus is preceded one clock period earlier by a "re pulse" upon the "read-pulse bus." These transfers are ma possible by the continuous application of the p_c, p_a, p_s and pulses, respectively.

To accept a number from the digit bus, the shift register sequenced as follows: At a clock-pulse time the first digit established in the storage element A0, as shown in Fig. 2. T next quarter pulse transfers the content of A0 to S0, the cont of A1 to S1, and so on up to A40 and S40. The following pu of the p_1 group, which is generated in the associated bina counter, transfers the content of S0 to A1, the content of to A2, and so on up to S39 and A40. After 40 of these operatio the first digit appears in S40, the second in S39, etc., and fortieth in S1.

Reading from the register which is initiated by a read pu requires the 40 pulses of the p_2 group to be applied to the ment S40, in addition to those pulses which are required dur the writing process. These pulses transfer the contents of S to the digit bus at clock-pulse times.

As shown in Fig. 2, the p_1 pulse group is counted by me of the binary counter, which is initiated by either a read pu or a write pulse. The 40 pulses of the p_2 group which required only during the reading process are obtained by us each of the p_1 pulses to gate out the following clock pulse fr the continuously-generated clock-pulse train.

Each of the six stages of the binary counter consists of a sim ring-of-two counter from which every second input pulse gated through to the following stage. Counting is initiated either a read pulse or a write pulse, which first sets the cou to zero and then admits shift pulses to the input of the cou by opening the input gate. When 40 pulses have been coun a set of six "coincidence" gates permits the passage of a cl pulse, which closes the input gate to the counter and thus minates the transfer operation of the register. In the exar chosen it is seen that the number leaving the register during reading process re-enters the register via the element A0. this way the number is retained in the register. During writing process, however, the incoming number replaces old one.

The unit described consists of a shift register, a binary cou and some associated circuits. The binary-gating element

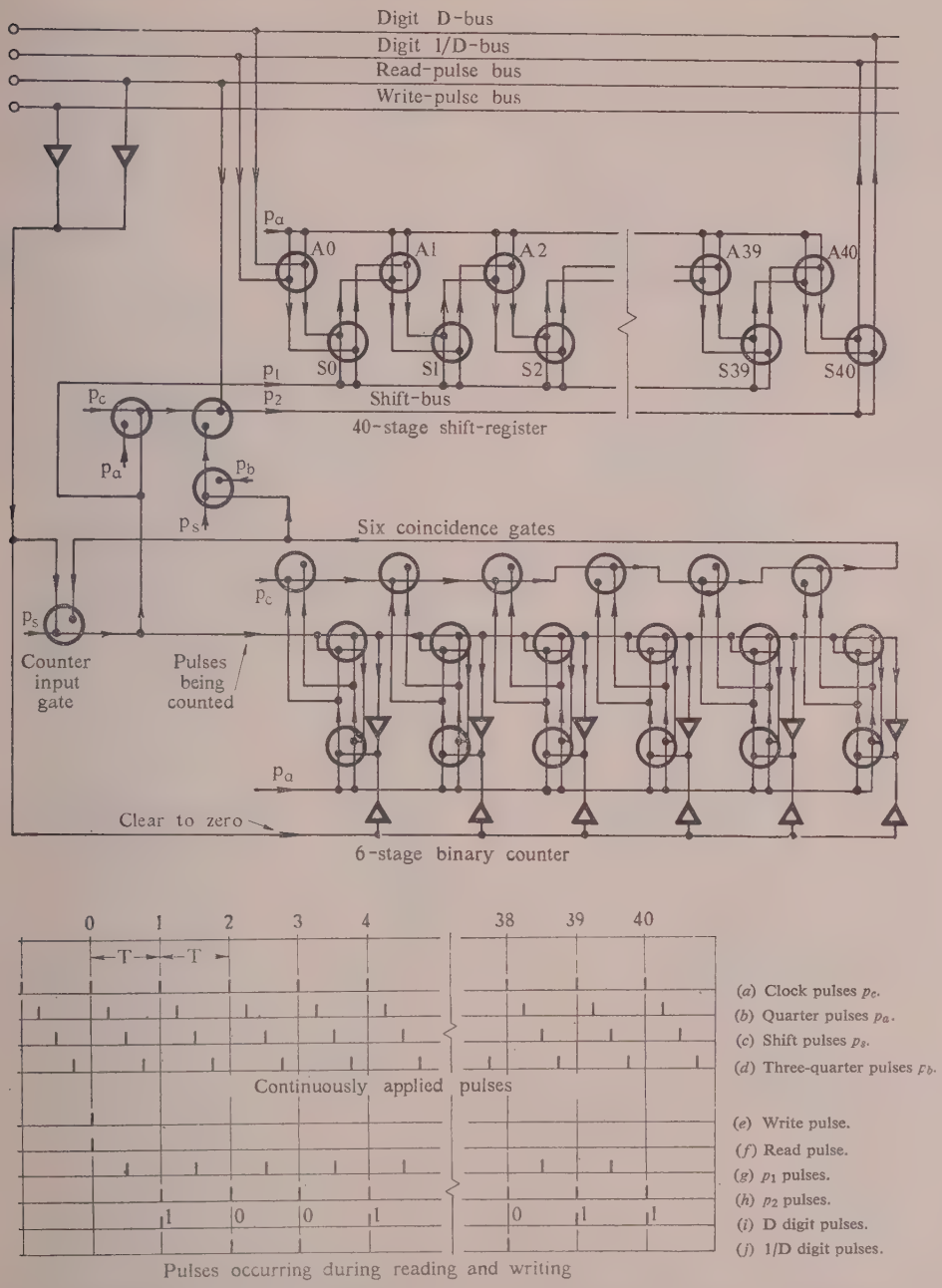


Fig. 2.—A 40-stage shift register and associated binary counter, assembled from binary-gating elements.

ilarly be used in the construction of binary adders, multi-
ers, decoders, ring counters, etc. There is no reason to
pect that the applications are limited to these cases—in fact,
se examples confirm the building-block nature of the binary-
ing element.

(3) BINARY-GATING TUBE

(3.1) Electrical Principles

The bistable section of the binary-gating tube is based on the nciple suggested by Sharpless[6] and used recently by Hollway[4] counter tubes. As indicated in Fig. 3, the bistable section consists of an electron gun, two pairs of deflector plates, and a pair of symmetrically located targets. Each target is connected to a load resistor and to the opposite deflector of the second pair. Such an arrangement may possess two electrical stable states, which are distinguished by the position of the electron beam. The beam may strike either target, and in doing so be held there by the deflection forces resulting from the feedback connections between targets and deflectors. The state of the device is indicated electrically by the target potentials, and it is from these targets that connections are made to the gate sections of the tube. The device may be set to a required state by momen-

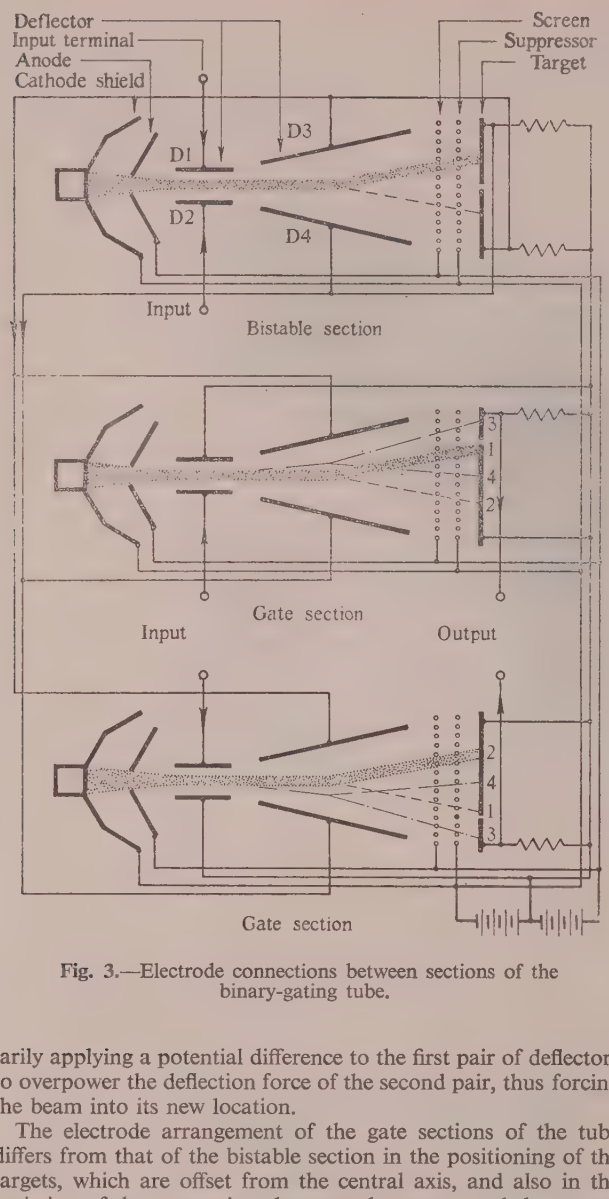


Fig. 3.—Electrode connections between sections of the binary-gating tube.

tarily applying a potential difference to the first pair of deflectors to overpower the deflection force of the second pair, thus forcing the beam into its new location.

The electrode arrangement of the gate sections of the tube differs from that of the bistable section in the positioning of the targets, which are offset from the central axis, and also in the omission of the connections between the targets and the second pair of deflectors. Connections are made between targets of the bistable section and the second pair of deflectors of the gate section, so that in the absence of a potential difference between the first pair of deflector plates of the gate section, their electron-beam location is similar to that of the beam in the bistable section. These locations are indicated in the diagram by target positions 1 and 2, respectively, target position 1 being associated with the "gate open" condition and target position 2 with the "gate closed" condition. Information appearing as a voltage pulse upon the input deflector of a gate section momentarily causes the beam to move either from target location 1 to 3 or from location 2 to 4 depending upon the state of the bistable section. In the former case, which corresponds to the gate-open condition, the beam strikes the output target, resulting in an output signal across the load resistor. In the latter case the beam deflection is insufficient to move the beam on to the output target, and no signal is produced at the output.

The binary-gating tube has been designed to operate with anode and screen grid connected to a 400-volt supply. Equa satisfactory performance is observed when this voltage lies the range 300–500 volts. The deflectors and targets are p ferably operated at about half the anode voltage, i.e. 200 vo Target currents observed with the anode at 400 volts are appro mately 0·2 mA. The pulses applied to the trigger deflectors the bistable section and to the input deflectors of the g sections are 25 volts in amplitude, although in a well-alig tube this value may lie in the range 15–35 volts. From a c sideration of the beam current and trigger-pulse amplitu together with the requirement that the output of one tube m be connected directly to the input of another, it follows that target load resistance required is 125 kilohms. In the bista section of the tube the relatively small external stray capacita present at the targets enables the sensitivity of the deflect which cross-connect to the targets to be made only half that the trigger deflectors, without a loss of overall switching spe The load resistances used in the bistable section are consequen 250 kilohms.

(3.2) Construction

As indicated in Fig. 4, the binary-gating tube is similar general appearance and form of construction to the more c ventional tubes. The plane-shaped electrodes are moun between mica discs to provide mechanical support and electri insulation. Significant features of the tube are the electron g the deflection system and the target assembly.

In the electron gun three ribbon-like electron beams formed, each originating from a separate portion of the comm indirectly-heated cathode, and emanating from one of th apertures in the anode of the gun. Beam formation takes pl in the region between the cathode shield and the anode, wh electrons from the cathode are confined to a wedge-shap region. Upon leaving the aperture the electrons move in a be through the deflection system to a line focus at the targets.

Between the electron gun and the targets the path of the be electrons is controlled by two pairs of deflectors. The f pairs of deflectors associated with each beam are electrica separate from one another and are mechanically supported glass beads. The common connections required between second pairs of deflectors enable the construction to be simplif by using two long flat plates to form a pair of deflectors comm to all three beams. In addition to controlling the gen direction of the electron paths, the deflection system is ma tained at a lower potential than that of the anode and scr grid to assist in the focusing of the beam. This also provides suppression of the secondary electrons produced at both anode aperture and the screen grid.

The target assembly consists of several flat plates suppo on glass beads. Secondary emission from the target surface suppressed by a closely spaced suppressor grid placed betw the screen grid and the targets and held at cathode poten The screen grid, placed between the suppressor grid and deflector system, maintains the electron velocities in this re and prevents the target potentials from influencing the elec paths.

The complete assembly is connected to a 12-pin button and enclosed in a cylindrical glass envelope to form a approximately 28 mm in diameter and 65 mm in length.

(3.3) Electron Optical Design

The use of the Pierce-type electron gun, in cases where it be applied, enables a first-order approximation of the elect geometry to be made by computational methods, thus confi trial-and-error experimental techniques to final adjustmer

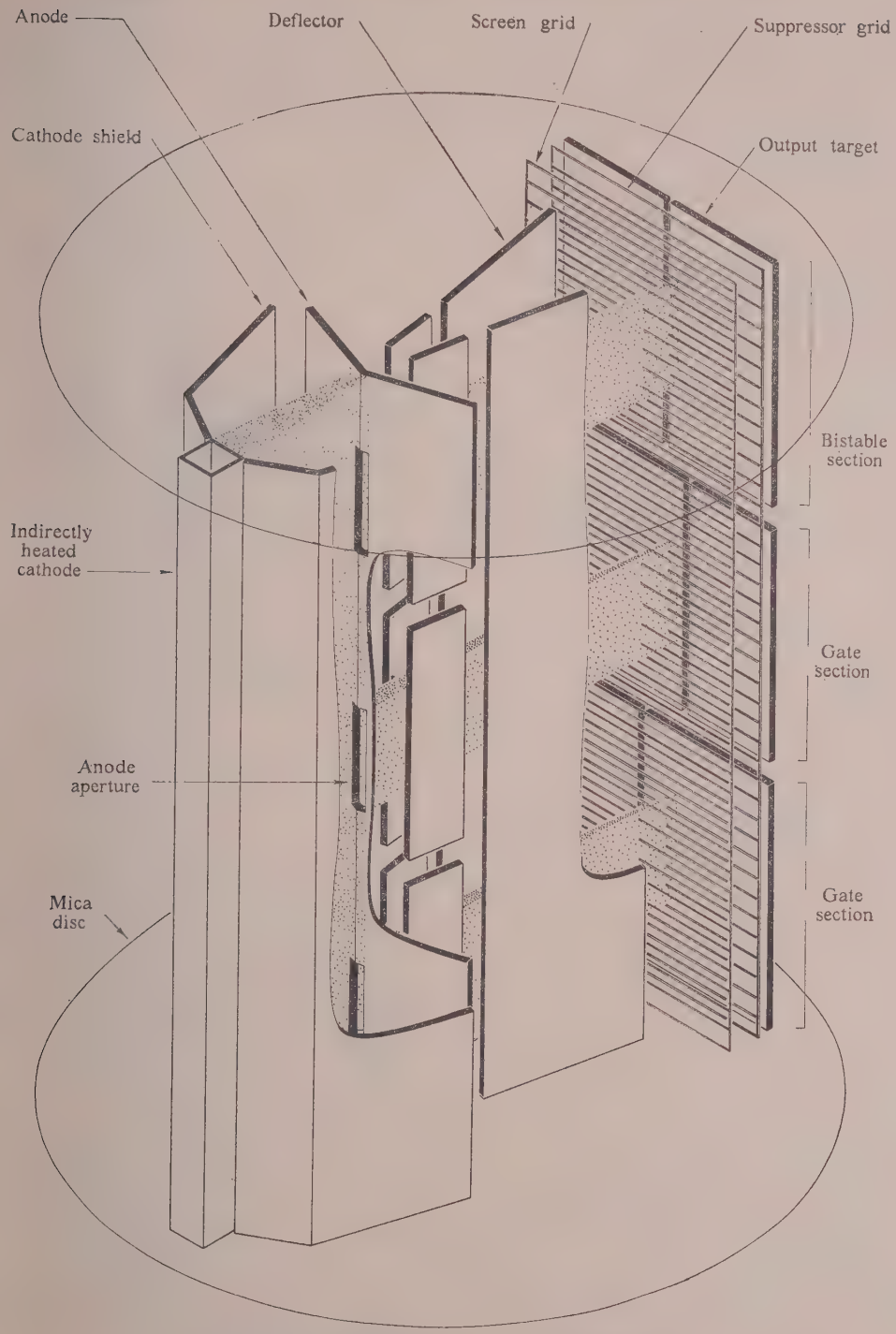


Fig. 4.—Electrode configuration of the binary-gating tube.

cond-order effects. The choice of the Pierce gun depends the predominance of space-charge over thermal-velocity in determining electron paths. In the binary-gating tube, ions are intermediate between these two cases, space predominating within the gun and thermal velocities in am beyond the gun. The gun to be described is designed y on the basis of space charge within both the gun and the beam, due consideration being given to the lens action of the anode aperture. Finally the width of the focus under these conditions is estimated by considering the beam spreading due to thermal effects.

Considering space charge only, Pierce[9] computed the approximate shapes of the cathode shield and anode necessary to confine the electrons in the gun to a wedge-shaped region. These curved

electrodes are difficult to make with accuracy, so for use in the binary-gating tube the simplified shapes of Fig. 4 have been obtained. These have been computed using a conduction-sheet analogue device, the necessary shapes being obtained by trial to give a close approximation to the Langmuir[10] diode-distribution relation along the boundary between the space-charge and the space-charge-free regions of the gun.

This potential relation is given by

$$V = k_1(r\beta^2)^{2/3} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where V = Field potential at radius r.

r = Radial position measured from the centre of curvature of the cathode.

$\beta = u - 2u^2/5 + 11u^3/120 - 47u^4/3\,300 + \ldots$

$u = \ln(r/r_c)$

r_c = Radius of curvature of the cathode.

k_1 = Constant.

Considering the lens action of the anode aperture, but neglecting both space-charge and thermal-velocity effects in the region beyond the anode, the electrons are directed to a cross-over point which is further removed from the anode than the "cathode-centre," as shown in Fig. 5. The focal length of the

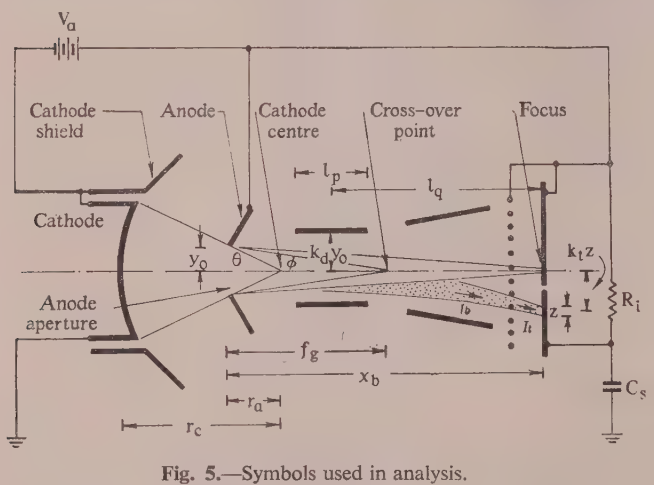


Fig. 5.—Symbols used in analysis.

divergent lens at the anode aperture is computed from the Davisson and Calbick[11] relation,

$$f_s = -2V_a/E_a \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where f_s = Anode-aperture focal length.

V_a = Field potential at the anode aperture.

E_a = Change in gradient through the aperture.

For the present analysis E_a is given the value of the derivative of the Langmuir potential relation at $r = r_a$. From the lens formula the focal length of the gun, f_g, is given by

$$\frac{1}{f_g} = \frac{1}{f_s} + \frac{1}{r_a} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Thus electrons leaving the anode aperture move towards the cross-over point and not the cathode centre.

Disregarding the effects of thermal velocities in the beam between the anode and the output targets, the effect of space charge is shown by Thompson and Headrick[12] to cause the beam to spread, and the focus to recede to a position further from the anode than the cross-over point. The binary-ga tube is designed for the limiting focusing condition, in which focus occurs in a distance of twice that of the cross-over p from the anode aperture, and the paths of the electrons at focus are tangential to the axis. The beam perveance for case is given by

$$\frac{I_b}{l_c V_a^{3/2}} = \frac{41{\cdot}8 \times 10^{-6} y_0}{x_b^2} \quad . \quad . \quad . \quad .$$

where, I_b = Total beam current.

l_c = Length of cathode.

y_0 = Half the anode-aperture width.

x_b = Beam length measured from the anode apertu

The limiting focusing condition is obtained when the follow two conditions are fulfilled:

(*a*) The gun perveance, as computed by the Langmu cylindrical-diode current relation, is equal to the beam pervea required by the Thompson and Headrick beam space-cha spreading relation for the limiting focusing condition:

$$\frac{I_b}{l_c V_a^{3/2}} = \frac{14{\cdot}66 \times 10^{-6}\theta}{r_a\beta^2\pi} = \frac{41{\cdot}8 \times 10^{-6} r_a \theta}{x_b^2}$$

where $\theta = y_0/r_a$

Therefore $r_a^2\beta^2 = 0{\cdot}111 x_b^2$ . . . . . .

(*b*) The paths of electrons leaving the anode aperture directed towards a cross-over point located midway betw the anode aperture and the required focus point:

$$f_g = x_b/2 \quad . \quad . \quad . \quad . \quad . \quad .$$

By trial solution of eqns. (5) and (6), the necessary proportions are

$$r_c = 0{\cdot}59\, x_b \quad . \quad . \quad . \quad . \quad .$$

$$r_a = 0{\cdot}21\, x_b \quad . \quad . \quad . \quad . \quad .$$

In practice, it is found that the focus may be improved, and current picked up by the deflectors reduced, if a limiting aper is included whose width is one-third of the anode aperture, which is placed at a distance of three times the anode-aper width from the anode. For clarity this aperture is not sh in Fig. 4. However, its effect must be taken into account w computing the beam current, which in this case is reduce one-third of its original value. In the binary-gating described

$$x_b = 13\text{mm}$$

From eqns. (7*a*) and (7*b*)

$$r_c = 7{\cdot}7\text{mm}$$

$$r_a = 2{\cdot}7\text{mm}$$

Allowing for the effect of the limiting aperture

$$y_0 = 0{\cdot}075\text{mm}$$

$$l_c = 4\text{mm}$$

$$V_a = 400 \text{ volts.}$$

Therefore, from eqn. (4),

$$I_b = 0{\cdot}59 \text{ mA.}$$

By experiment the screen grid was found to reduce the current at the target to 0·20 mA.

The important assumption in the space-charge-spre theory is that the electron beam at the anode aperture is homogeneous and homocentric; these being conditions p

he Pierce-type electron gun when thermal velocities are egarded. Langmuir[13] and Pierce[14] have shown that the ct of thermal velocities is to limit the current densities which be obtained in electron beams, and to destroy the fine-line s which would otherwise be possible in ribbon-like beams n aberrationless electron optical system.

the absence of space charge in the beam the focus occurs ne cross-over point, as shown in Fig. 5. The equation for effects of thermal velocities in ribbon beams shows that the imum current density which can be obtained at the focus t is given approximately by

$$J_{max} = J_0 \frac{2}{\sqrt{\pi}} (eV/kT)^{\frac{1}{2}} \sin \phi \quad . \quad . \quad . \quad . \quad (8)$$

re, J_0 = Cathode current density.
V = Beam potential.
ϕ = Beam convergence angle at the cross-over point for electrons leaving the anode with zero velocity.

ratio of the beam width at the cross-over point to that at cathode is given by J_{max}/J_0. When $V = 400$ volts and the h of the cathode is 0·054in, the beam width at the cross- point is 0·024in, as compared with 0·018in at the anode ture. Assuming the beam to continue spreading at the e rate, the expected focus width at the target would be 0in. The formula shows that the use of a limiting aperture ot be expected to reduce the spreading due to thermal cities to any appreciable extent. It can be expected to cept and remove the electrons of high thermal energy and reduce the stray current entering the deflector electrodes.

he equation for the effect of thermal velocities shows that maximum current density which can be obtained in a beam creased as the angle of convergence of the electron paths is eased. Thus the beam focus can be improved by using a lensing lens. In the binary-gating tube, such a focusing t is obtained by operating the deflection system at a lower ntial than the anode and the screen grid, the latter two rodes being electrically connected. The operation of the in this manner does not prevent the output of one tube g connected directly to the input of another, since the ence of the screen grid enables the output targets of the to be connected through their appropriate load resistors he same supply as the deflectors. In fact, with the tube ribed the best operating conditions were found when the et and deflectors were operated at about half the potential he anode and screen, as indicated in Fig. 3. The focusing t of this lens may be computed by using one or other of the ous numerical methods, such as are given by Klemperer Wright[15] or Gans,[16] based upon the differential equation notion of the paraxial electrons. Alternately, as has been e in the design of this tube, the potential of the deflection em necessary to give the best line focus may be obtained by riment. It was found that the line-focus width could be ced from 0·030in to 0·010in by lowering the deflector ntials to one-half that of the anode.

econdary electrons produced at the target are suppressed in binary-gating tube by means of a suppressor grid placed veen the screen grid and the targets. The method used for nating the dimensions of these grids is similar to that used he design of pentodes, and presented in a convenient form Spangenberg.[17]

erce[18] had given an expression for the limiting trans- ductance of beam-deflection amplifier tubes in terms of mal velocities and the capacitance between the deflector es by combining the expression for the maximum current density with that for the deflection sensitivity. For use in the design of the binary-gating tube, an expression is derived for the time-constant associated with the change of potential of a target when the beam is suddenly switched on to it, in terms of the width of the focus at the target, the beam deflection, the total stray capacitance, the anode potential, and the ratio of the deflector-plate separation to the beam width at the anode. The thermal velocities are not considered directly in this derivation, because space-charge and focusing effects in the deflection system cannot be neglected in this case. However, their effect is implied in the term for the beam width at the focus. The variation of the capacitance between the deflector plates as a function of their separation is not taken into account, since their capacitance is comparable with that of the stray capacitance to earth of the wiring when the tube is used in computer circuits.

In the circuit shown in Fig. 5, the following symbols are used:

$2y_0$ = Beam width at limiting aperture.
$2y_0k_d$ = Separation of the deflector plates.
l_c = Length of the cathode (beam thickness).
x_b = Length of the beam.
z = Beam width at target.
V_a = Anode potential.
I_b = Beam current.
I_t = Current entering the target.
R_1 = Load resistance.
C_s = Stray capacitance at the target.
V_t = Input, or output, signal (the signals are of equal amplitude to enable direct connection to be made between tubes).

The target response to an input signal which causes the beam to be suddenly switched on to it is given by

$$v_0 = I_tR_1[1 - \exp(-t/T)] \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where
$$T = C_sR_1 = C_sV_t/I_t \quad . \quad . \quad . \quad . \quad . \quad (9a)$$

In the deflection system

$$V_t = 4V_a \frac{k_dy_0}{l_p} \frac{k_tz}{l_q} \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where, l_p = Length of the deflector plates.
l_q = Distance from the centre of the deflector plates to the target.

From eqn. (4),

$$I_t = k_sI_b = 41{\cdot}8 \times 10^{-6} V_a^{3/2} \frac{k_sy_0l_c}{x_b^2} \quad . \quad . \quad (11)$$

where k_s = Fraction of the beam current passing through the screen grid and striking the target.

By substitution in eqn. (9*a*),

$$T = 9{\cdot}6 \times 10^4 \frac{1}{V_a^{\frac{1}{2}}} \frac{k_d}{k_s} \frac{k_tzx_b^2}{l_pl_ql_c} C_s \quad . \quad . \quad . \quad (12)$$

In the binary-gating tube described

$C_s = 15\ \mu\mu$F (approx.)
$l_p = 2{\cdot}25$mm
$l_q = 10{\cdot}5$mm
$k_d = 6{\cdot}3$
$k_tz = 1{\cdot}0$mm
$k_s = 0{\cdot}34$
$l_c = 4{\cdot}0$mm
$V_a = 400$ volts
$x_b = 13$mm

Substituting these values in eqn. (12), $T = 2{\cdot}4$microsec. This agrees with the time-constants associated with the waveforms in Fig. 6 and with the observed maximum speed of operation of 150 000 changes of state per second.

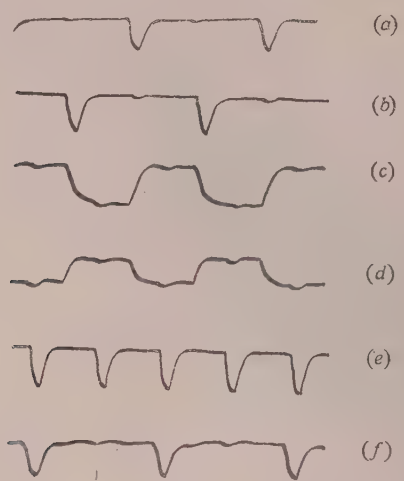


Fig. 6.—Oscillograms showing the performance of the binary-gating tube at a frequency of 50 000 pulses/sec.

(*a*) Pulses applied to trigger deflector D1.
(*b*) Pulses applied to trigger deflector D2.
(*c*) Waveform at bistable deflector D3.
(*d*) Waveform at bistable deflector D4.
(*e*) Input pulses applied to gates "a" and "b."
(*f*) Output pulses from gate "b."
(*g*) Output pulses from gate "a."

(4) CONCLUSION

Approaching the design of digital systems from the viewpoint of basic functions has led to the establishment of three basic elements and the practical realization of two of them in the form of a beam-deflection vacuum tube. Experience may show that the binary-gating tube does not represent the best possible compromise between complexity of the tube and that of the overall system. A simpler element which can perform either bistable or gate functions by a simple change of external connections may lead to a more economical overall arrangement. Whatever tube is finally chosen for the purpose, it does appear that its use would simplify the design, construction and maintenance of digital machinery by reducing the number and variety of its components.

(5) ACKNOWLEDGMENTS

The author acknowledges the guidance and encouragement of Professor D. M. Myers of the Electrical Engineering Department of the University of Sydney.

Acknowledgments are also due to Mr. W. R. Blunden of the Mathematical Instruments Section of the Commonwealth Scientific and Industrial Research Organization, for facilitating laboratory work; to Mr. L. G. Bellamy of the Mathematical Instruments Section for skilful assistance in the construction experimental tubes; to Mr. R. E. Aitchison of the Electri Engineering Department of the University of Sydney, a Mr. D. L. Hollway of the Division of Electrotechnology C.S.I.R.O., for valuable discussions on vacuum work; to Mr. James and the Staff of the Valve Laboratory of C.S.I.R.O.; a to Dr. D. Gabor of the Electrical Engineering Department the Imperial College of Science and Technology for advice the preparation of the paper.

The assistance received by way of grants from the N Zealand Council for Scientific and Industrial Research un the National Research Scholarship Regulations is a acknowledged.

(6) REFERENCES

(1) HAEFF, A. V.: "The Memory Tube and its Applications Electronic Computation," *Mathematical Tables a other Aids to Computation*, 1948, **3**, p. 281.
(2) RAJCHMAN, J. H.: "The Selectron—a Tube for Select Electrostatic Storage," *ibid.*, 1947, **2**, p. 359.
(3) WILLIAMS, F. C., and KILBURN, T.: "A Storage System Use with Binary-Digital Computing Machines," *P ceedings I.E.E.*, Paper No. 763, November, 1948 (Part II, p. 81).
(4) HOLLWAY, D. L.: "An Electron-Beam Decimal Coun Tube," *Nature*, 1950, **165**, p. 856.
(5) RAJCHMAN, J. H.: "Final Report on Computron," *R Report.*
(6) SHARPLESS, J. K.: "Theory and Techniques for the Des of Electronic Digital Computers," *Moore School Electrical Engineering*, 1946, **2**, Lecture 16.
(7) JONKER, J. L. H.: "Valves with Ribbon-Shaped Electr Beams," *Philips Research Reports*, 1950, **5**, p. 6.
(8) JONKER, J. L. H., and VAN GELDER, Z.: "New Electro Tubes employed as Switches in Communication En neering," *Philips Technical Review*, 1951, **13**, p. 49.
(9) PIERCE, J. R.: "Rectilinear Electron Flow in Beam *Journal of Applied Physics*, 1940, **11**, p. 548.
(10) LANGMUIR, I., and BLODGETT, K. B.: "Currents Limi by Space-Charge between Coaxial Cylinders," *Physi Review*, 1923, **22**, p. 347.
(11) DAVISSON, C. J., and CALBICK, C. J.: *ibid.*, 1932, **42**, p. 5
(12) THOMPSON, B. J., and HEADRICK, L. B.: "Space-Cha Limitations on the Focus of Electron Beams," *P ceedings of the Institute of Radio Engineers*, 1940, p. 318.
(13) LANGMUIR, D. B.: "Theoretical Limitations of Catho Ray Tubes," *ibid.*, 1937, **25**, p. 977.
(14) PIERCE, J. R.: "Limiting Current Densities in Electr Beams," *Journal of Applied Physics*, 1939, **10**, p. 715.
(15) KLEMPERER, O., and WRIGHT, W. D.: *Proceedings of Physical Society*, 1939, **51**, Part II, p. 296.
(16) GANS, R.: *Zeitschrift für Technische Physik*, 1937, p. 41.
(17) SPANGENBERG, K. R.: "Vacuum Tubes" (McGraw-H 1948).
(18) PIERCE, J. R.: "Theoretical Limitations to the Tra conductance in Certain Types of Vacuum Tub *Proceedings of the Institute of Radio Engineers*, 1943, p. 657.

85.832: 621.374.32

Monograph No. 108 M
Sept. 1954

A BEAM-DEFLECTION VALVE FOR USE IN DIGITAL COMPUTING CIRCUITS

By M. W. ALLEN, B.E., Student.

(*The paper was first received 5th April, and in revised form 9th June,* 1954. *It was published as an* INSTITUTION MONOGRAPH *in September,* 1954.)

SUMMARY

fferent methods used in the logical design of computing circuits ummarized, and from these, desirable properties are suggested for versal computing element.
description is given of a ribbon-beam-deflection valve based on suggestions. Methods of application are illustrated with ts and oscillograms using a single valve alternatively for binary ion, binary subtraction, or control of an output gate by various vibrator circuits.

(1) INTRODUCTION

has been appreciated for some time that the conventional valve is not ideal for switching circuits, since simple opera- often require several valves. Consequently digital com- s use large numbers of valves and components, which usly affect their reliability. Various proposals have been for universal switching elements in an attempt to reduce otal amount of equipment and the variety of circuits. A ground to the idea of universal elements is provided by the us symbols and methods used in logical design. These will iefly reviewed.

system based on the analogy to the nervous system was oped by von Neumann[2] and others from the work of Pitts McCulloch, who proposed a description of the nervous m based upon "two indication elements." These elements, n they called neurons, have the following properties: each on has an output line and a number of input lines. The lines carry stimuli to excite or inhibit a signal on the ut line (the response being immediate). All the signals are lued, i.e. of the on-off type and not continuous. An "all one" output response changes when the number of simulta- s inputs exceeds a certain number, called the threshold of euron. Some of the symbols used and their characteristics hown in Fig. 1. The number in the circle is the threshold he neuron; an inhibit stimulus overrides all excitation li.

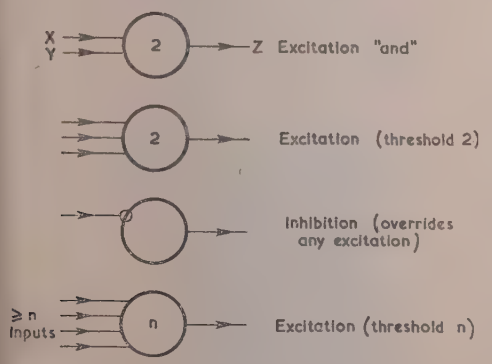


Fig. 1.—Neuron symbols.

Different connections of these elements can be written, using the rules of Boolean algebra. A difficulty arises in practice from the fact that there is very little correspondence between these symbols and the appropriate vacuum-tube circuits. Several modified schemes have been proposed to overcome this difficulty.

One method[4] uses special symbols rather than neurons to represent the valve circuits, and associated with each symbol is an algebraic operator; also included are Tables which list the different switching combinations of up to four variables in terms of these operators. (The switching combinations represent all the possible arrangements for four bi-valued inputs and a single bi-valued output). The operator expressions can be converted to symbolic form from which detailed circuits can be drawn. These Tables show that considerably more vacuum tubes are required than neurons for a given switching function. A number of other similar schemes have also been developed.

There appear to be two alternative solutions to simplify circuits. The first is to use only a single basic element, different combinations of this being used to derive the other elements. Several devices have been proposed on this basis. Speedy[3] has described a special valve and circuit which combines the properties of a trigger circuit and two electronic gates, and has illustrated how this may be used as a building block.

Another proposal[1] is the universal decision elements shown in Fig. 2. The switching Table is written as a form of matrix.

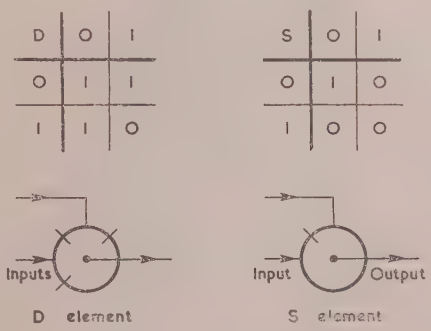


Fig. 2.—Decision elements.

The two input signals, each having possible values of 0 or 1, form the upper and left-hand side, and the output from the element, which can likewise be 0 or 1, is given by the intersection. In the symbol which is used to represent this matrix, a short line through the circumference indicates that the output is a 1 in the corresponding quadrant of the matrix. The universal elements are the D and S elements. One reason for the choice of these simple elements with two inputs and one output is that they are easier to analyse. More complex elements having more than two inputs and more than one output, although unwieldly in logical systems, may be useful in practice.

espondence on Monographs is invited for consideration with a view to tion.
Allen is in the Mathematical Instruments Section, C.S.I.R.O., Australia.

The following are suggested as desirable properties in a device to be used in a somewhat similar way to neurons:

(*a*) Four inputs, two of each polarity, to be used as excitation and inhibit connections. These inputs should preferably be of equal sensitivity, and such that direct links can be made between elements.
(*b*) Output terminals to give positive or negative signals.
(*c*) Complete isolation between inputs, and between inputs and outputs, so that they do not interact with each other.
(*d*) A power gain sufficient to drive as many as three other elements in parallel.

The present paper describes a ribbon-beam-deflection valve consisting of two interconnected elements, each having the above properties. The reason for this particular choice is that it enables binary addition, subtraction and other operations of interest to be carried out in a single valve which is limited to a 12-pin base. The application of the valve is illustrated with circuits and oscillograms.

It is interesting to observe that a still more useful element would be one whose proportions were capable of simple dynamic control. An example of such an element would be one which could be changed from addition to subtraction on application of a control signal derived from the previous operation, or from an external source. Similar elements are useful in logical systems, and would be equally useful in automatic computing. At present they seem difficult to realize practically.

(2) DESCRIPTION OF VALVE

Fig. 3 shows the essential features of one of the valves used. The valve is approximately 2¼in long and 1¼in in diameter. It

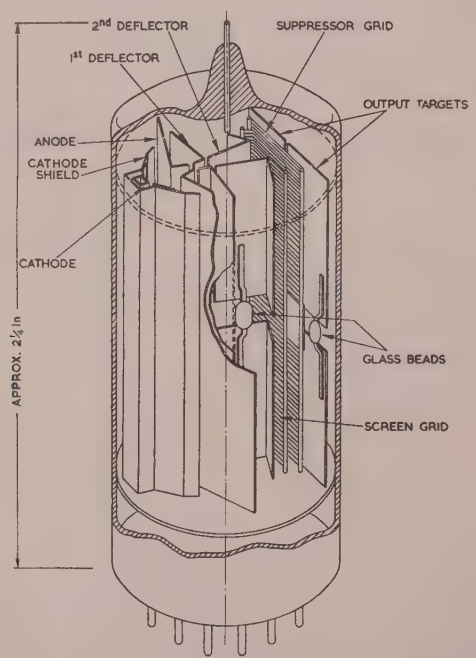


Fig. 3.—Valve details.

consists of a simple electron gun producing a ribbon-type beam which is directed axially outwards but deflected by sets of deflectors through screen and suppressor grids on to collecting targets. The direct voltage of the deflectors is half that of the anode producing a focusing field and assisting the repulsion of secondary electrons from the anode. The second deflectors divided so that they can act independently on their respec sections of the beam. The beam is divided into sections be reaching the targets by a strip across the suppressor g Recent investigations have shown that considerable impr ments would result from a back-to-back arrangement of sect using two sides of a common cathode. This would be preferred arrangement for future work.

The essential elements of each section for design purposes the arrangements of targets and deflectors. Each section h divided target and two pairs of deflectors (the first pai deflectors being common to both sections). The bottom of targets is divided centrally, and for the top pair the divi is offset, so that with all the deflectors at the same potential, beam divides centrally between the targets in the bottom sect and lies wholly on one target in the top section. The sensiti of all the deflectors is the same, and their d.c. level is the s as that of the targets, so that direct links can be made betw targets and deflectors. These properties simplify cir arrangements.

The input signals applied to the deflectors are norm negative voltage pulses. The output signal is then taken corresponding negative voltage pulse from a load resi attached to the target on which the beam impinges. In s cases positive voltage pulses are used, and these cases are tinguished by a primed letter. A corresponding positive ou pulse occurs in a load resistor attached to the target from wh the beam shifts on application of the input. A unit input p is considered as one which when applied to a deflector will s the beam at the target a distance somewhat greater than beam width.

Using the notation of Fig. 4 for the various deflectors targets, it may be seen that the following equations relate output target currents to the deflector-voltage inputs.

For the bottom section of the valve

$$i_{B1} = i, \text{ and } i_{B2} = 0$$

when $$D_2 - D_1 + D_{4B} - D_{3B} \geqslant \tfrac{1}{2} \quad . \quad . \quad . \quad .$$

and for the top section $i_{T1} = i$, and $i_{T2} = 0$

when $$D_2 - D_1 + D_{4T} - D_{3T} \geqslant 1 \quad . \quad . \quad . \quad .$$

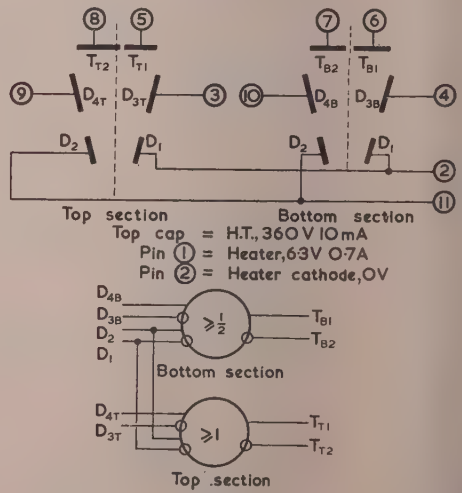


Fig. 4.—Base connections and symbol.

he relations may be summarized in the symbols of Fig. 4. nections to deflectors which produce deflection of the beam ne direction are shown plain, while connections to deflectors h produce deflection in the opposite direction are indicated small circle surrounding the connection. Plain and circled ections to the targets are used to distinguish those parts on hich the electron beam is directed by negative pulses applied milarly marked deflectors.

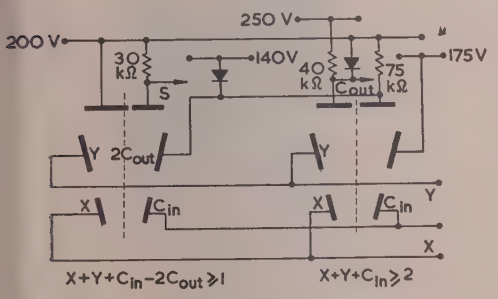
ne numbers in the circle correspond to those in eqns. (1) (2) and specify the conditions for current in the targets. number is determined by the target offset, but it may be ed by a constant voltage difference on any deflector pair; by applying an appropriate bias to D_{3B} the number $\geqslant \frac{1}{2}$ be altered to $\geqslant 0$, $\geqslant 1$, $\geqslant 2$, etc.

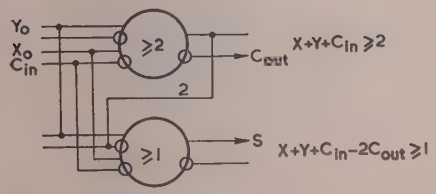
(3) CIRCUIT APPLICATIONS

(3.1) Binary Addition

forming the sum of two numbers, each step consists of the tion of the addend and augend, together with the carry digit the previous step.

n = Base of the number system.
d y = Addend and augend digits.
C_{in} = Carry digit from the previous step.
S = Sum-digit output.
C_{out} = Carry-digit output.





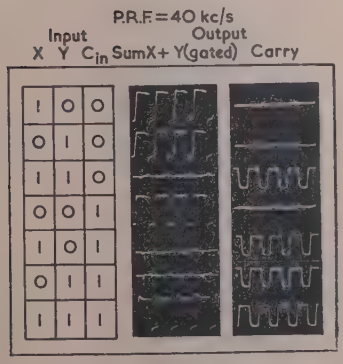


Fig. 5.—Binary addition.

then the rules for the output sum and carry digits may be stated as follows:

C_{out} occurs when

$$x + y + C_{in} \geqslant n \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and S occurs when

$$x + y + C_{in} - nC_{out} \geqslant 1 = S \quad . \quad . \quad . \quad (4)$$

In the binary system, $n = 2$, and we have

$$C_{out} = 1, \text{ when } x + y + C_{in} \geqslant 2$$

and $$s = 1, \text{ when } x + y + C_{in} - 2C_{out} = 1$$

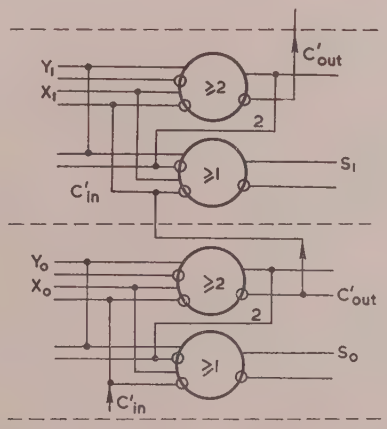
These relations are similar to those given earlier relating target currents and deflector voltages. Fig. 5 shows these relations in symbolic form, and also the detailed circuit. The $\geqslant 2$ relation for the carry digit is obtained by a lower d.c. level on D_{3B}. In the circuit shown

x and y are supplied as negative voltage pulses of unit amplitude.
C'_{in} is a positive voltage pulse of unit amplitude.
$2C_{out}$ is obtained by using a load resistance sufficient to produce a voltage pulse of at least 2 units, clipped by a diode to reduce it accurately to 2 units.
C'_{out} is a positive voltage pulse output.

It is apparent that spurious results can occur during the rise and fall of the various inputs. This does not necessarily cause complications in practice, since the results are always read into a storage register whose input gates may be closed during the transient period. In obtaining the waveforms shown the sum digits have been gated in this way. No spurious results can occur with the carry digits.

In the case of parallel addition, it is necessary only to connect the C'_{out} digit of one stage to the C'_{in} of the next.

Fig. 6 shows a two-digit parallel adder, and the waveforms illustrate some of the more interesting combinations involving a carry between stages.



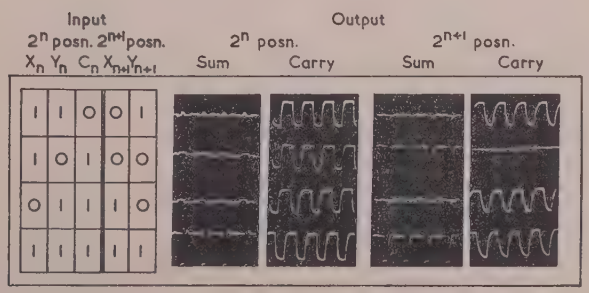


Fig. 6.—Parallel binary addition.

For a complete serial adder the C'_{out} digits must be delayed one digit period to provide the C_{in} digit for the following step. Fig. 7 shows a circuit which was operated at 120 kc/s with 5 microsec pulses. An amplifying valve is included in the $2C_{out}$ link to reduce transients, and a second valve feeds an 8-microsec delay line for the carry digits which are shaped by diodes to provide C'_{in}.

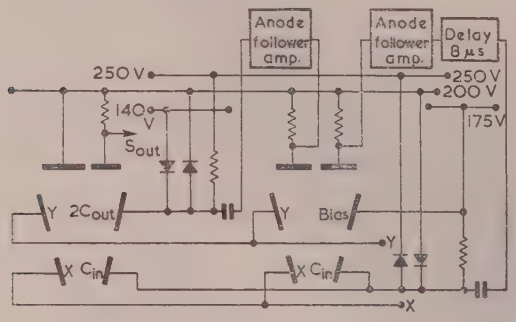


Fig. 7.—Serial binary addition.

(3.2) Binary Subtraction

Each step in the subtraction of the two numbers consists of subtraction from the minuend of the digits of the subtrahend, together with the borrow from the previous step, proceeding always from the least significant digit. The rules for binary arithmetic are as follows:

$$\text{Borrow } B_{out} = 1 \text{ when}$$
$$x - y - B_{in} \leqslant 1 \quad . \quad . \quad . \quad . \quad . \quad (5)$$
$$\text{Difference } D = 1 \text{ when}$$
$$x - y - B_{in} + 2B_{out} = 1 \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Fig. 8 shows the symbolic form and the detailed circuit. The $\leqslant -1$ relation is achieved by a bias on D_{4B}.

x, y, and B_{in} are supplied as negative voltage pulses of unit amplitude.
$2B_{out}$ is obtained using resistances and a clipping diode.
B_{out} is taken from a dividing resistance.

As before, transient errors can occur during the rise and fall of the pulses, and these have been gated out.

Parallel and serial subtraction circuits can be made in a similar way to the corresponding addition circuits.

It is apparent that this method of circuit design may be applied to other switching circuits. The procedure is to examine the switching Table for relations between the input and output, of the following type:

$$D_1 - D_2 + D_{3B} - D_{4B} \geqslant \text{or} \leqslant n,$$
$$D_1 - D_2 + D_{3T} - D_{4T} \geqslant \text{or} \leqslant n_2$$

The requirements for binary addition and subtraction have resulted in the following design limits for the valve:

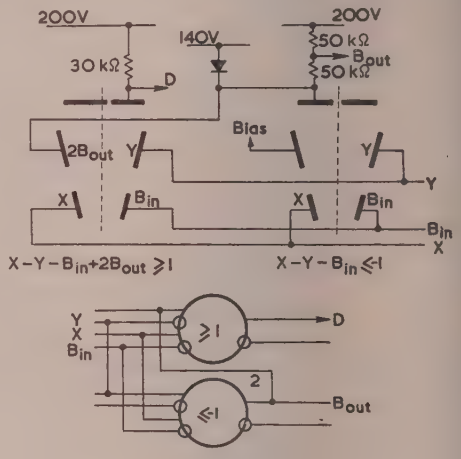
$$3 \geqslant n \geqslant -3$$
$$2 \geqslant D_1 - D_2 \geqslant -2$$

The relations should be checked to see that these limits are not exceeded.

The circuits which follow involve the control of a gate by a second element, and use the same valve as for binary addition and subtraction. They have been selected as the ones of most interest in digital computing.

(3.3) Bistable Storage

The application of the binary gating valve as a building b for switching circuits has been described by Speedy.[3] The section valve may be used similarly, except that only one ga available. The circuit for a bistable section and coupled



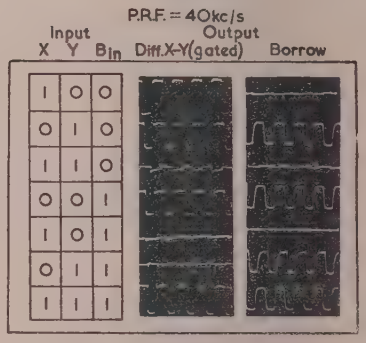


Fig. 8.—Binary subtraction.

is shown in Fig. 9. The symmetrically located targets are c connected to the common pair of deflectors, and to the su voltage through load resistors. This results in two stable st which may be switched by pulses on the second pair of deflec

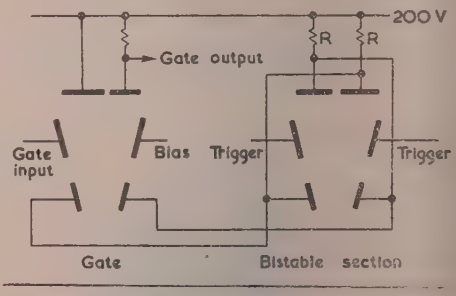


Fig. 9.—Bistable storage with an output gate.

The remaining section is then used as a gate coupled b common pair of deflectors. The gate target is offset mec cally, and a bias is applied to D_{3B}, so that in one state the of the beam lies just off the target, and in the other it lies be one and two beam widths from the target. A unit pulse gate shifts the beam towards the gate target, so that the can strike the output target in one state and not in the oth

ig. 10 shows a modification of this circuit to provide an ıt as well as an output gate. This circuit would be of use in ic registers in a parallel machine.

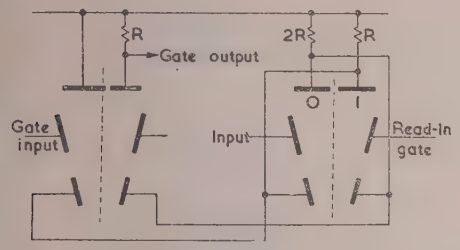


Fig. 10.—Register storage.

he load resistances for the symmetrical targets have been sen as $2R$ and R, so that the second stable position is shifted beam width further off centre. The state can only be tched by the coincidence of a digit pulse on D_{4B} and a positive pulse on D_{3B}. The read-out gate operates as before.

(3.4) Binary Counter

he action of the deflectors in a beam valve is in many ects equivalent to that of a grid in a triode. Thus circuits be derived directly from equivalent grid-valve circuits. This been done in the circuit shown in Fig. 11.

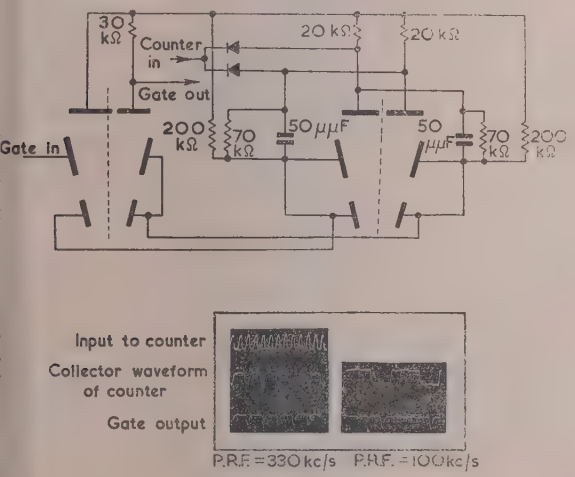


Fig. 11.—Binary counter with gate.

he direct link between target and deflectors is replaced by a stance-capacitance network, and negative input pulses fed ugh diodes. The load resistances and the resistances of the s-over network are chosen so that there are two stable es for the beam. A sharp negative pulse applied to the input le causes the target at the high potential (no beam current) ollow the input, but transmits nothing to the other low-potential target. Because of the cross-coupling capacitor, the fraction of this which appears at the deflectors exceeds the steady-state value required for bistable action (which is determined by the resistances alone), and is sufficient to change the state. The following pulse switches the beam back, and so on.

In grid-tube circuits the switching speed is reduced by the loss in the step-down networks necessary because of the difference in d.c. level between the grids and anodes. In the beam valve the d.c. levels are the same, and the network loss can be minimized. This is the reason why beam valves with a much lower equivalent mutual conductance (in terms of milliamperes per volt on deflector or grid) perform comparably in such circuits. The waveforms of Fig. 11 illustrate the operation of the counter and coupled gate. Input pulses of about 2 microsec in width are applied to the counter, and identical pulses occurring midway between them are applied to the gate. Operation is shown at 100 kc/s and 330 kc/s, the latter being the limiting frequency.

(3.5) Multivibrators

It is apparent that, with the diodes separated, the circuit of Fig. 11 acts as a bistable multivibrator.

The bistable multivibrator can be converted to a monostable or astable multivibrator in exactly the same way as in grid-valve circuits. A monostable multivibrator is obtained by eliminating one resistor coupling a target to a deflector, but leaving in the capacitor. The beam can then remain in one stable state as before, but can only remain in the other state for a time dependent on the time for the discharge of the cross-coupling capacitor. One of the diodes can be omitted. Similarly, by eliminating both cross-coupling resistors, the circuit becomes equivalent to a free-running or a stable multivibrator. The remaining section of the valve can then be used as a coupled gate.

(4) ACKNOWLEDGMENTS

The author wishes to acknowledge the contributions of Professor D. M. Myers of the Electrical Engineering Department, University of Sydney, under whose guidance the investigation was undertaken; Dr. C. B. Speedy, formerly of the Electrical Engineering Department, University of Sydney, whose development of "binary gating valve" led to this investigation; and Mr. L. G. Bellamy and members of the C.S.I.R.O. Vacuum Laboratory for construction of the experimental valves. The work described is part of the research programme of the Mathematical Instruments Section of C.S.I.R.O.

(5) REFERENCES

(1) Goodell, J. F.: "The Foundations of Computing Machinery," *Journal of Computing Systems*, 1952, 1, No. 1.
(2) Hartree, D. R.: "Calculating Instruments and Machines" (University of Illinois Press, 1949).
(3) Speedy, C. B.: "The Function of Basic Elements in Digital Systems," (see page 49).
(4) "Synthesis of Electronic Computing and Control Circuits" (Annals of the Computation Laboratory of Harvard University 1951, 27).

VIBRATORY POWER CONVERTORS: AN ANALYSIS OF PERFORMANCE AND DESIGN

By R. H. EVANS, B.Sc., Graduate.

(*The paper was first received* 21*st August*, 1953, *and in revised form* 12*th May*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in September*, 1954.)

SUMMARY

The modes of operation of the various circuits employed in vibratory convertors are briefly described and a more detailed analysis is then made. Formulae are derived for use in design and in most instances these are presented graphically.

The paper includes calculation of the transformer flux density, the buffer capacitance, and the no-load (constant) losses, together with load characteristics for a.c. outputs with resistive or inductive load, and d.c. outputs employing full-wave, half-wave or voltage-doubling rectifying circuits. The ripple voltage across d.c. outputs is calculated and its harmonic content analysed. Formulae for input current, current in the transformer windings, and copper loss are also included, and are utilized to compare the power ratings for a given size of transformer operating in any of the various circuits.

The paper deals mainly with conventional circuits for low or medium power, up to about 50 watts; however, one section analyses some special circuits for d.c. outputs of high power in which the current pulses in the transformer are so shaped that the vibrator contacts open at virtually no load.

LIST OF PRINCIPAL SYMBOLS

A_i = Net cross-sectional iron area of core, cm^2.
A_k = Net cross-sectional copper area of winding, cm^2.
$\hat{B}$ = Peak flux density in core, gauss.
B_c = Instantaneous flux density at end of a closed interval of the contacts, gauss.
B_m = Average flux density during an open interval of the contacts, gauss.
C = Reservoir capacitance, farads.
C_b = Buffer capacitance, farads.
C_s = Smoothing capacitance, farads.
f = Frequency of vibrator reed, c/s.
g = R.M.S. current factor for inductive a.c. load. (This is the factor by which $V/2\pi fL_T$ is multiplied to obtain the load current I_{2R}.)
H_m = Magnetizing force corresponding to B_m, ampere-turns/cm.
I_1 = Mean battery current (load component only), amperes.
I_2 = Mean output current, amperes.
I_{2R} = R.M.S. output current from transformer, amperes.
I_m = Magnetizing current corresponding to B_m, amperes.
I_n = Mean battery current at no load, amperes.
I_{w1} = R.M.S. current in primary winding, amperes.
I_{w2} = R.M.S. current in secondary winding, amperes.
i_c = Instantaneous current at end of closed interval of contacts, amperes.
i_o = Instantaneous current at end of open interval of contacts, amperes.
J = Current density in transformer windings, amp/cm^2.
k_1 = Closure-time ratio of an interrupter contact, defined as the fraction of a complete cycle during which the contact is closed.
k_2 = Closure-time ratio of a rectifying contact.
L = Transformer leakage inductance, henrys, referred to secondary winding. (For d.c. outputs it is the t series inductance up to the reservoir capacitor, comprises the leakage inductance together with series inductance added to achieve contact unloadi
L_2 = Inductance of load, henrys.
L_n = Inductance of transformer primary winding at no l henrys.
L_s = Inductance of smoothing choke, henrys.
$L_T = L + L_2$.
l = Mean length of magnetic path in transformer core, c
l_g = Effective length of air-gap in magnetic path, cm.
M = Weight of iron in core, g.
N = Number of turns in half primary winding, or in w primary if not centre tapped.
n = Transformer turns ratio. [This is the ratio (secon turns)/N, or (turns in half secondary)/N if the secon winding is centre-tapped.]
P_1 = Input power to supply load and copper loss, watts.
P_b = Power loss due to over-buffering, watts.
P_k = Copper loss in transformer and circuit, watts.
q = Order of harmonic (e.g. for second harmonic, $q = 2$
R = Circuit resistance referred to the secondary winc ohms.
R_1 = Resistance of primary circuit, ohms.
R_2 = Load resistance, ohms.
R_e = Effective resistance of convertor viewed from ou terminals, ohms.
R_n = Resistance representing the iron losses in the equiva circuit of the transformer, ohms.
R_s = Resistance of the winding of the smoothing choke, oh
$R_T = R + R_2$.
r = Ripple factor. The ripple factor for any ripple monic is the factor by which $\hat{V}_r$ is multiplied to ob the r.m.s. value of the harmonic.
s = Resistance multiplying factor (see Section 5.1).
t_c = Time of one closed interval of the contacts = seconds.
t_o = Time of one open interval of the contacts, seconds.
V = Peak induced secondary e.m.f. at no load = nV_1, vo
V_1 = Battery voltage.
V_2 = Mean rectified output voltage.
V_{2R} = R.M.S. output voltage.
$\hat{V}_r$ = Peak-to-peak ripple voltage.
v_c = Instantaneous voltage across the reservoir capacit the end of a charging interval.
v_o = Instantaneous voltage across the reservoir capacit the end of a discharging interval.
δ = Density of iron, g/cm^3.
λ = Coefficient expressing the relative permissible p ratings of transformers.
$\hat{\Phi}$ = Peak flux, corresponding to $\hat{B}$, maxwells.
Φ_c = Flux corresponding to B_c, maxwells.

Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Evans is at the Royal Aircraft Establishment, Farnborough.

(1) INTRODUCTION

(1.1) General

The vibratory convertor is one means of converting power in form of direct current such as the 24-volt d.c. supply of an craft, to power in the form of alternating current, or direct rent at a different voltage. The Joint Service "Guide on rators"[1] describes the general principles and gives practical ails concerning the choice of circuit and the selection or ign of individual components, including the vibrator and the nsformer. Various other general accounts[2,3] have been blished, while Mitchell[4] has surveyed recent trends in develop-nt and Allen[5] has investigated the special problem of contact sion in vibrators.

The present paper provides a more detailed analysis of the ious circuits and of the transformer operating conditions, l includes the derivation of design formulae. Only the steady-te conditions are dealt with and not the initial switching nsients. Where possible, the design equations are converted dimensionless form and displayed graphically. The ranges of merical values included for the operating characteristics of the rator, such as frequency and closure-time ratio, are those en in the relevant Joint Service Specifications; they will, how-r, also be applicable to the majority of vibrators manu-tured for the general market.

(1.2) Mode of Operation

A vibratory convertor consists essentially of a vibrator, a nsformer and various other components for waveform cor-tion, smoothing and radio interference suppression. The d.c. rce from which it operates is usually of low voltage and low pedance, such as an accumulator. The vibrator itself behaves a two-way switch, maintained in continuous vibration by ctromagnetic means.

The primary winding of the transformer is usually centre-ped, the centre tap being connected to the positive side of supply [Fig. 1(*a*)]. The vibrator reed is connected to the gative side of the supply, and the fixed contacts on either side the reed are joined to the outer ends of the primary winding. us current flows alternately in the two half windings in posite directions, an alternating flux is produced in the core, l hence an alternating voltage is induced in the secondary nding. During that part of the cycle in which the contacts closed, the constant battery voltage is applied to the primary nding and the voltage waveform is therefore flat-topped. The veform of the secondary voltage is similar, since substantially same flux links both windings.

The spacing of the contacts is so adjusted that each is closed about 0·39 of a cycle; this ratio is termed the closure-time o. A buffer capacitor C_b is connected across one of the ndings, in order to maintain the magnetizing current while the ntacts are open. If it were absent the rapid collapse of gnetizing current as the contacts opened would induce high tages in the windings, probably resulting in destructive arcing he contacts.

f a d.c. output is required, a metal or valve rectifier [Figs.) and 1(*c*)] may be employed, but the most efficient method is use a self-rectifying vibrator. This has an additional set of ntacts operating in conjunction with the reed, and if these ntacts are connected to the outer ends of a centre-tapped ondary winding, a unidirectional but intermittent voltage btained between the centre tap and the reed [Fig. 1(*d*)]. This tage is smoothed by the reservoir capacitor C_r, and a further cuit consisting of the choke L_s and capacitor C_s. The closure-e ratio of each rectifying contact is usually about 0·34.

The split-reed type of vibrator has two reeds, rigidly connected mechanically but insulated electrically, so that the output may be insulated from the input, as in Fig. 1(*e*). Other applications of this vibrator are the voltage-doubling output [Fig. 1(*f*)] and the double-pole change-over input [Fig. 1(*g*)].

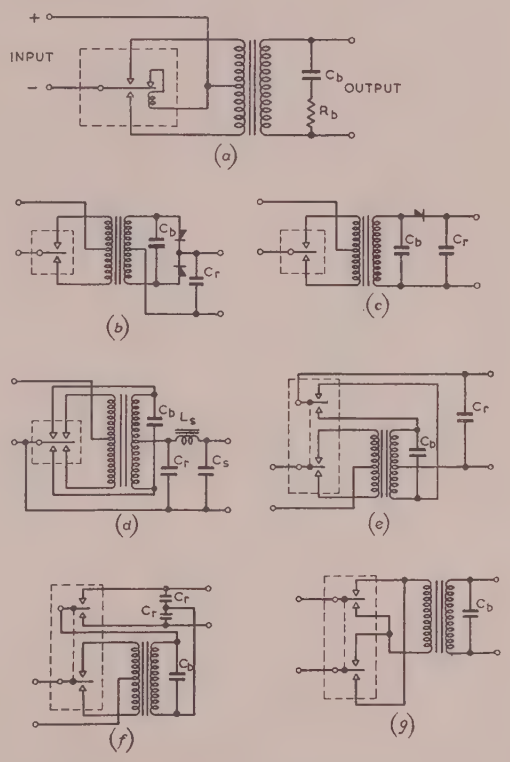


Fig. 1.—Vibratory-convertor circuits.

(*a*) Basic circuit, a.c. output.
(*b*) Interrupter vibrator with separate rectifier, full-wave.
(*c*) As (*b*) but half-wave rectification.
(*d*) Self-rectifying vibrator, full-wave.
(*e*) Split-reed self-rectifying vibrator, full-wave.
(*f*) Split-reed, self-rectifying vibrator, voltage-doubling circuit.
(*g*) Double-pole change-over input, a.c. output.

(2) TRANSFORMER FLUX

(2.1) Flux Density and Number of Primary Turns

The primary winding is designed so that the flux density in the core at no load at the highest input voltage likely to be encountered has an acceptable value. This value is lower than in mains-operated transformers in order to minimize the peak current on starting, which may cause destructive arcing at the vibrator contacts. For ordinary silicon-iron laminations, typical values of the peak working flux density $\hat{B}$ are 8 000 gauss for inputs of 12 volts or below, and about 5 700 gauss for higher input voltages, where the danger of sustained arcing is greater.

The voltage across each half of the primary winding at no load is trapezium shaped and of peak value equal to the battery voltage V_1 [Fig. 2(*a*)]. The resulting flux is shown in Fig. 2(*b*), the shaded portion being that maintained by the current flowing from the buffer capacitor. If it is assumed to a first approximation that the sloping sides of the voltage waveform are linear, the average value over a half-cycle is $(1 + 2k_1)V_1/2$ volts, and at no load this may be equated to the average induced e.m.f.,

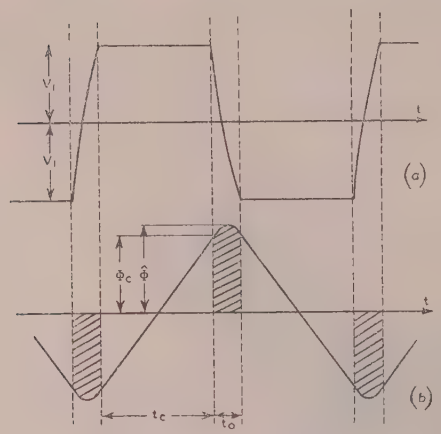


Fig. 2.—No-load waveforms.
(*a*) Voltage across half primary winding.
(*b*) Core flux.
The contacts are closed during the intervals t_c, open during t_o.

given by $4f\hat{\Phi}N \times 10^{-8}$ volts. Hence the number of turns required in each half of the primary winding is given by

$$N = \frac{(1 + 2k_1)V_1 \times 10^8}{8f\hat{B}A_i}$$

$$= \frac{0{\cdot}89V_1 \times 10^8}{4f\hat{B}A_i} \quad \text{for } k_1 = 0{\cdot}39 \quad . \quad . \quad (1)$$

Comparison may be made with the expression for a sinusoidal supply, namely

$$N = \frac{V_1 \times 10^8}{4{\cdot}44f\hat{B}A_i} = \frac{0{\cdot}90V_1 \times 10^8}{4f\hat{B}A_i} \quad . \quad . \quad . \quad (2)$$

The two expressions are almost identical, but it should be noted that V_1 in the vibrator formula is the peak value (the battery voltage) not the r.m.s. value.

Sometimes[1] the value of flux density B_c attained at the end of a closed interval of the contacts, rather than the true peak value $\hat{B}$, is regarded as the important design figure. Since a steady e.m.f. of V_1 volts is induced by a flux change of $2\Phi_c$ in a time k_1/f seconds, the expression for B_c is

$$B_c = \frac{k_1V_1 \times 10^8}{2fA_iN} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

In one method of calculating the buffer capacitance (Section 3.2) a knowledge of the average flux density B_m during an open interval of the contacts is required. By integration over this interval its value may be shown to be

$$B_m = \frac{(1 + 4k_1)V_1 \times 10^8}{12fA_iN} \quad . \quad . \quad . \quad . \quad (4)$$

For $k_1 = 0{\cdot}39$ the three flux densities are related by $B_c = 0{\cdot}875\hat{B}$ and $B_m = 0{\cdot}96\hat{B}$.

(2.2) Asymmetry of Flux Waveform

Unsatisfactory operation of vibrator convertors may be caused by conditions in which the transformer flux has a unidirectional component. For example, the employment of half-wave rectification [Fig. 1(*c*)] would produce such a component and this circuit is therefore not recommended, except possibly as an auxiliary supplying only a fraction of the total output. A similar flux unbalance would result from the adoption of h wave transformer input, that is a circuit in which the prim winding has no centre tap and the current flowing is me interrupted instead of being periodically reversed as in conventional arrangement.

Assymmetry of the flux waveform may also result if the c tacts on either side of the vibrating reed have unequal con resistances or unequal closure-time, or if the halves of transformer primary winding have unequal resistances. N mally, the effect of these inequalities is small, but it incre with load and may be important in vibratory convertors for h powers.

(3) BUFFER CAPACITANCE AND NO-LOAD LOSSES

(3.1) Calculation of the Buffer Capacitance by means of the Equivalent Circuit

(3.1.1) Theoretically Ideal Buffering.

Under no-load conditions the equivalent circuit is that Fig. 3, where the transformer primary winding is represented an inductance L_n carrying the magnetizing current and parallel resistance R_n carrying the loss current. The effect leakage inductance and the resistance of the primary wind

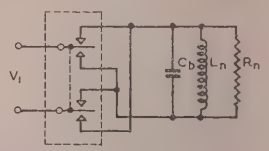


Fig. 3.—Equivalent circuit referred to the primary winding at no load.

may be neglected at no load. For simplicity the buffer capac C_b is shown connected across the primary winding an double-pole change-over input circuit is considered.

At the moment when the contacts open, the circuit is at voltage V_1, and it then oscillates freely until the contacts c after the interval t_o. Ideally, the capacitance should be related to the other circuit parameters and the intervals t_c t_o that the voltage just reaches the value $-V_1$ as the cont close. The voltage then has the waveform shown in Fig. and the current in L_n has a shape similar to that of the flux shown in Fig. 2(*b*). If a capacitance larger than the idea employed, the transformer is said to be over-buffered, and too small a value it is under-buffered. It is usual in prac to adopt over-buffering, but the theoretical ideal will be c sidered first.

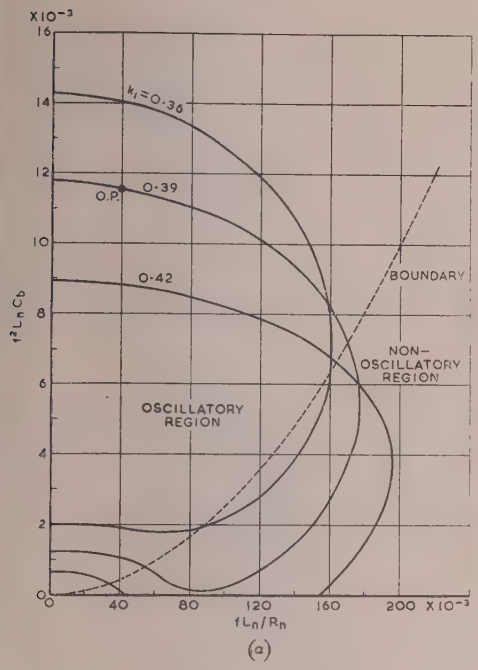
It may be shown by the method outlined in Section 11.1 for ideal buffering the circuit parameters should be related by equation

$$\frac{(\alpha^2 + \omega^2)t_c \sin \omega t_o}{2\omega} - \cos \omega t_o - \cosh \alpha t_o = 0 \quad .$$

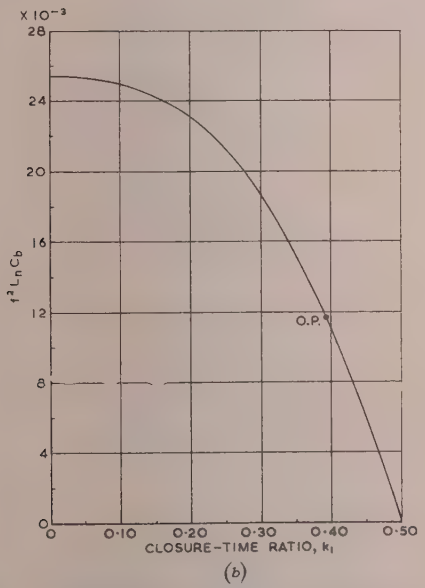
where $\alpha = -1/2C_bR_n$ and $\omega^2 = 1/L_nC_b - (1/2C_bR_n)^2$. T is a corresponding expression for the non-oscillatory condit when $1/L_nC_b < (1/2C_bR_n)^2$.

When the substitutions $t_c = k_1/f$ and $t_o = (1 - 2k_1)/2f$ made, the equation is seen to be in three variables, nam k_1, $f^2L_nC_b$ and fL_n/R_n. The relationship between the last tw plotted in Fig. 4(*a*) for three values of k_1 centred on 0·39. boundary between the oscillatory and non-oscillatory regio also shown.

It will be observed that in general there are two value $f^2L_nC_b$ for any value of fL_n/R_n, but it is found that the l



(a)



(b)

Fig. 4.—Relation between circuit parameters of Fig. 3 for ideal buffering.

(*a*) General.
(*b*) With negligible iron loss ($R_n \rightarrow \infty$).
O.P. Typical practical operating point.

ue is not of practical importance because its adoption causes voltage to pass through a peak before reaching the final ired value of $-V_1$; also the peak current is high.

ig. 4(*a*) also indicates that for a given value of k_1 there is a ximum value of fL_n/R_n, above which there is no corresponding ue of $f^2L_nC_b$. This means that if the losses are sufficiently h (R_n small) it is impossible to choose a capacitor to give the al waveform. The waveform then resembles that for the r-buffered condition [Fig. 5(*e*)].

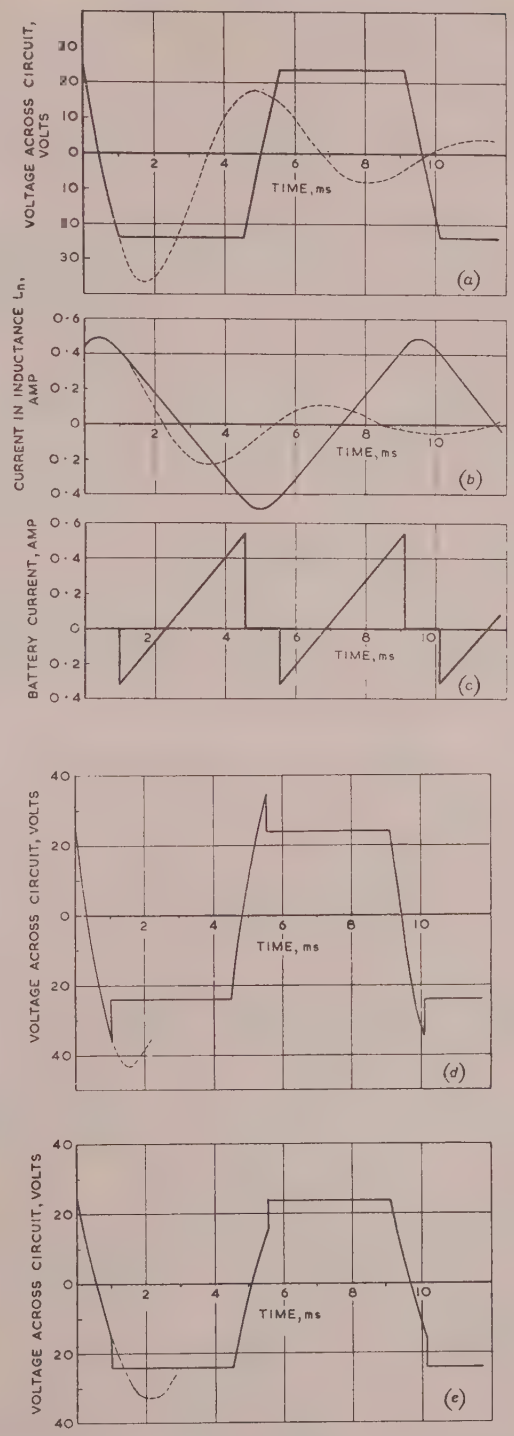


Fig. 5.—Calculated voltage and current waveforms at no load.

——— Actual waveform.
- - - - - Free oscillation.

(*a*)–(*c*) With ideal buffering, $C_b = 9{\cdot}45\ \mu$F.
(*d*) With under-buffering, $C_b = 8\ \mu$F.
(*e*) With over-buffering, $C_b = 12\ \mu$F.
$V_1 = 24$ volts, $f = 110$ c/s, $k_1 = {\cdot}39$, $L_n = 0{\cdot}1$ henry, $R_n = 225$ ohms.

At the opposite extreme is the case of negligible losses ($R_n \to \infty$) for which eqn. (5) reduces to

$$\tfrac{1}{2}\omega t_c = \cot \tfrac{1}{2}\omega t_o \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where $\omega^2 = 1/L_nC_b$. This equation is in two variables, k_1 and $f^2L_nC_b$, and is plotted in Fig. 4(*b*). A typical value of fL_n/R_n is 0·04, and since in this region the curves of Fig. 4(*a*) are fairly flat it is seen that neglect of the iron losses is usually justified.

(3.1.2) The Effect of Buffer Capacitance upon Voltage Waveform.

The effect of buffer capacitance upon waveform is illustrated in Fig. 5. The moment when the contacts open has been chosen as the time origin; the figure shows what the waveforms would be if the contacts did not close again (damped oscillations) and also the waveforms which are actually obtained owing to the cyclic switching of the contacts.

Figs. 5(*a*)–5(*c*) relate to the ideal value of buffer capacitance given by the appropriate curve of Fig. 4(*a*) and were plotted with the aid of eqns. (36), (37), (38) and (40).

Figs. 5(*d*) and 5(*e*) were plotted with the aid of eqns. (37), (38) and (39). In the under-buffered condition the voltage falls suddenly from 34·6 volts to 24 volts as the contacts close, and in the over-buffered condition it rises from 15·5 volts to 24 volts.

Departure from ideal buffering involves a waste of power and is a source of radio interference, but under-buffering is the more destructive condition owing to the peak voltages which occur. In order to allow for the fall in closure-time during the life of the contacts it is common practice to over-buffer initially to such an extent that the vertical portion of the waveform is about 40% of the peak-to-peak amplitude.

(3.2) Direct Calculation of the Buffer Capacitance from the Magnetization Curve of the Core Material

The method of Section 3.1.1 requires a knowledge of the inductance of the transformer primary winding, which would be calculated from the magnetization curve of the core material. However in practical design a more direct method, which neglects the effect of the iron losses, is to employ the curve to read off the magnetizing force H_m corresponding to the average flux density B_m existing during an open interval of the contacts [eqn. (4)]. The average magnetizing current, I_m, which is to be supplied by the buffer capacitor may thus be determined, due allowance being made for the effective length, l_g, of the air-gap introduced by the construction of the core.

For the purpose of calculation, it will be assumed that all the capacitance is supplied across one half of the primary winding, or across the whole primary in the case of a double-pole change-over input. The equivalent capacitance for any other winding may be derived, since it is inversely proportional to the square of the number of turns in the winding.

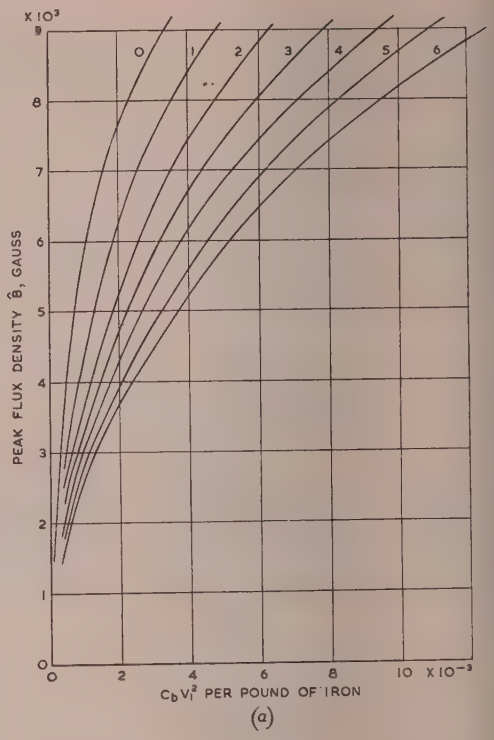
With the usual degree of over-buffering the free oscillation during an open interval of the contacts encompasses a voltage change of $1{\cdot}2V_1$, so that the average capacitor current in this interval is $1{\cdot}2V_1C_b/t_o$. Since $t_o = (1 - 2k_1)/2f$, the required capacitance is

$$C_b = \frac{(1 - 2k_1)I_m}{2{\cdot}4fV_1} \quad . \quad . \quad . \quad . \quad . \quad (7)$$

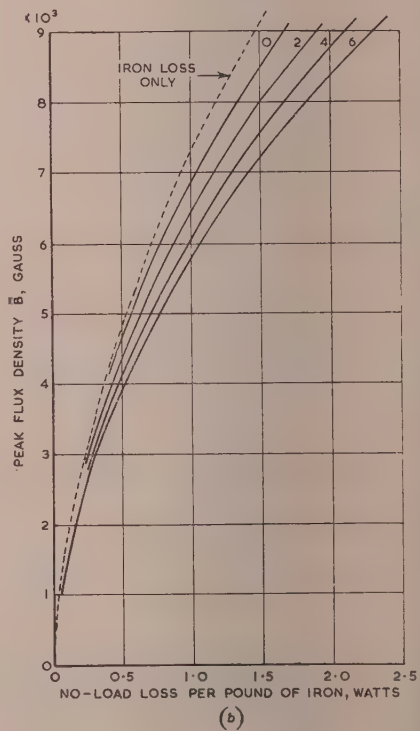
A more convenient form for expressing the result graphically in terms of the peak working flux density is obtained by substituting $I_m = (H_m l + 0{\cdot}8B_m l_g)/N$, and then eliminating N by means of eqn. (1). Finally, by writing lA_i as M/δ, and rearranging, the expression

$$\frac{C_bV_1^2}{M} = 3{\cdot}33\frac{1 - 2k_1}{1 + 2k_1}\frac{\hat{B}}{\delta}\left(H_m + 0{\cdot}8B_m\frac{l_g}{l}\right) \times 10^{-8} \quad . \quad (8)$$

is obtained.



(*a*)



(*b*)

Fig. 6.—Buffer capacitance and total no-load losses for transform incorporating typical silicon-iron laminations.

(*a*) C_b is recommended capacitance in farads, allowing for contact wear, and is referred to half-primary. Constructed for $k_1 = 0{\cdot}39$.
(*b*) Constructed for $k_1 = 0{\cdot}39$ and $f = 110$ c/s; laminations 0·014 in thick. Le to curves, multiplied by 10^{-4}, denote values of ratio $_g/l$.

Since for a given core material, H_m is a function of B_m, which turn is a known function of $\hat{B}$, the value of $C_b V_1^2$ per unit ight of iron depends only upon $\hat{B}$, k_1, and the ratio l_g/l. In rticular it is independent of frequency. Fig. 6(*a*) gives ues of $C_b V_1^2$ per pound of iron for a typical silicon iron, king k_1 to be 0·39. Once the size of lamination and depth of ck have been decided it is possible to determine the buffer pacitance, because the values of all the relevant parameters are en known.

Fig. 6(*a*) indicates that a high proportion of the buffer capaci- ce is required for the air-gap, so that in the interests of space onomy it is essential to assemble the laminations with great re. The value of l_g may then be taken to be about 0·003 cm.

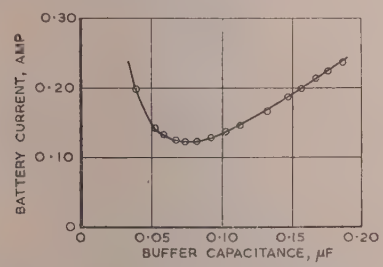
(3.3) Losses at No Load

3.1) Iron Loss.

A comparison may be made between the iron losses for a en core material, frequency and peak flux density when the x is produced by a trapezium-shaped voltage and a sinusoidal ltage respectively. It may be shown that the loss is lower vibrator operation, but the difference is so small that it is mmon practice to refer to curves of total iron loss which have en determined for sinusoidal waveforms.

3.2) Loss due to Over-Buffering.

It is found that if a vibratory convertor is operated at no d and the buffer capacitance is varied, there is a point at ich the battery current, and hence the input power, is a nimum; at this point practically all the input power is absorbed iron loss and the voltage waveform assumes very nearly the al trapezium shape. Experimental results for a typical nvertor are given in Fig. 7.



. 7.—Effect of buffer capacitance upon battery current at no load (experimental results).

Capacitor connected across secondary winding.
Peak induced e.m.f. = 206 volts.
Battery voltage = 12 volts.

The extra power loss with imperfect buffering is explained by fact that there is a sudden change of voltage across the pacitor as the contacts close, accompanied by a pulse of rent of high peak value, which gives rise to a loss in the istance of the transformer windings, battery leads, etc. With usual degree of over-buffering the change in the voltage oss one half of the primary winding is $0 \cdot 8 V_1$ volts, so that if capacitor is connected across this half-winding the energy s is $\frac{1}{2} C_b (0 \cdot 8 V_1)^2$. Since this loss occurs twice in each cycle, power loss is given by

$$P_b = 0 \cdot 64 f C_b V_1^2 \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Values of $C_b V_1^2$ against peak flux density have already been tted in Fig. 6(*a*) and thus for a given frequency only a change scale is required to convert this graph into one showing the r-buffering loss. An additional curve may be inserted giving iron loss for the particular core material and frequency, and hence, by addition, curves of total no-load loss may be plotted. Fig. 6(*b*) gives such curves for a typical silicon iron at 110 c/s. In typical cases the over-buffering loss is about 20% of the total no-load loss.

(4) A.C. OUTPUTS

(4.1) Resistive Load

(4.1.1) Load Characteristics.

At no load the voltage across each half of the primary winding is approximately trapezium-shaped and of peak value V_1, giving an r.m.s. value $\sqrt{(1 + 4k_1)} V_1/\sqrt{3}$. Thus the r.m.s. output is $\sqrt{(1 + 4k_1)} n V_1/\sqrt{3}$.

For low voltage inputs (e.g. 2 volts) the neglect of the voltage drop due to the no-load current flowing in the primary winding may not be justified, and in this case a corrected value (V_1') may be substituted for V_1 in all calculations. It is derived as follows. If the mean battery current at no load (given by the total no-load losses of the transformer) is I_n, then the mean value during a closed interval of the contacts is $I_n/2k_1$, so that the voltage transformed at no load is given by

$$V_1' = V_1 - R_1 I_n / 2k_1$$

where R_1 is the resistance of the primary circuit. This is the corrected value of input voltage and may be employed for any condition of load, and with rectified as well as a.c. outputs.

If a purely resistive load is connected across the secondary winding the voltage waveform will approach the rectangular shape of Fig. 8(*a*), because the buffer capacitor will be rapidly

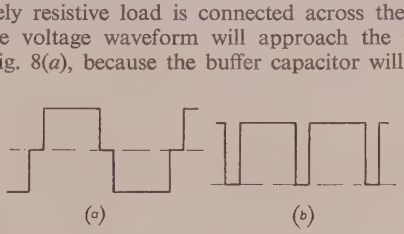


Fig. 8.—Waveforms with a.c. output, resistive load.
(*a*) Output voltage.
(*b*) Battery current.

discharged each time the contacts open; the output current will of course have the same waveform. If the r.m.s. output current is denoted by I_{2R}, the peak current is $I_{2R}/\sqrt{(2k_1)}$ and the peak output voltage is $nV_1 - RI_{2R}/\sqrt{(2k_1)}$. R is the total circuit resistance referred to the output side, and comprises the contact resistance together with the resistances of the half-primary and secondary windings. It follows that the r.m.s. output voltage is

$$V_{2R} = \sqrt{(2k_1)} n V_1 - R I_{2R} \quad . \quad . \quad . \quad . \quad (10)$$

so that the convertor behaves as a source of voltage $\sqrt{(2k_1)} n V_1$ in series with a resistance R. This characteristic is not valid near open-circuit, where, as mentioned above, the waveform is trapezium-shaped rather than rectangular.

The battery current is of the same waveform as the output current except that the rectangular pulses are unidirectional [Fig. 8(*b*)]. The average output current calculated over a half-cycle is $\sqrt{(2k_1)} I_{2R}$, so that the mean value of battery current is given by

$$I_1 = \sqrt{(2k_1)} n I_{2R} \quad . \quad . \quad . \quad . \quad . \quad (11)$$

This is the load component of the input current, supplying the copper losses (RI_{2R}^2) and the external load; to obtain the total current to the convertor, the components supplying the transformer constant losses and the vibrator driving power should be added.

Each half of the primary winding carries only alternate current

pulses, so that the r.m.s. current I_{w1} in the winding is $nI_{2R}/\sqrt{2}$. In terms of the battery current this may be expressed as

$$I_{w1} = I_1/2\sqrt{k_1} \quad . \quad . \quad . \quad . \quad . \quad (12)$$

(4.1.2) **Power Rating of the Transformer.**

It is well known that the power input at unity power factor to a transformer may be written

$$P_1 = \lambda f \hat{B} J A_i A_k \times 10^{-8} \quad . \quad . \quad . \quad . \quad (13)$$

For a sinusoidal input with a non-rectified output, λ has the value 2·22. To determine λ for a vibrator transformer with a centre-tapped primary winding and feeding a.c. into a resistive load we may rearrange eqn. (1) to make V_1 the subject and, multiplying by I_1, we obtain

$$P_1 = 8/(1 + 2k_1) f \hat{B} A_i N I_1 \times 10^{-8} \quad . \quad . \quad . \quad (14)$$

Now the primary and secondary windings occupy respectively net copper areas of $2I_{w1}N/J$ and $I_{w2}nN/J$, so that with the aid of eqns. (11) and (12) we obtain

$$I_1 N/\sqrt{k_1} + I_1 N/\sqrt{(2k_1)} = JA_k$$

Hence $$\lambda = \frac{8}{1 + 2k_1} \frac{\sqrt{(2k_1)}}{1 + \sqrt{2}} = 1{\cdot}64 \text{ for } k_1 = 0{\cdot}39$$

The coefficient λ may be employed as a measure of the permissible power rating of a transformer of given size. It is seen that the power for a given size of vibrator transformer would be 26% lower than that of a transformer with a sinusoidal input of the same frequency, even if the same flux density could be employed. Since the working flux density is normally considerably lower, the power rating is further reduced. It should be noted, however, that the vibrator frequency is usually at least 100 c/s, so that the rating would not necessarily be less than that of a 50 c/s mains transformer. Moreover, if a double-pole change-over circuit is adopted [Fig. 1(*g*)] the value of λ is improved to 1·98.

(4.2) Inductive Load

Sometimes a vibratory convertor is required to supply alternating current to a highly inductive load, such as a fluorescent lamp with its series stabilizing choke. In such an application, the buffer capacitance needed to supply the load current during the open intervals of the contacts is many times greater than that for normal applications, where only the transformer magnetizing current has to be supplied. In such highly inductive circuits, the voltage waveform—provided that the correct buffer capacitance is supplied—closely resembles the trapezium-shaped no-load waveform. The problem is to calculate the magnitude of the voltage of this waveform necessary to send a current of known r.m.s. value, I_{2R}, through a load consisting of a resistance R_2 and an inductance L_2 in series; and also to determine the optimum buffer capacitance under these conditions.

In the equivalent circuit shown in Fig. 9(*a*) the transformer resistance R has been added to the load resistance R_2 to give the total resistance R_T, and similarly the leakage inductance L has been added to L_2 to give L_T. All values are referred to the secondary; V is the peak induced secondary e.m.f. and is equal to nV_1. The problem is closely analogous to that of Section 3.1.1 (Fig. 3), the only difference being that the resistance and inductance are in series instead of in parallel. An outline of the solution is given in Section 11.2, where it is shown that the buffer capacitance should satisfy the following relation.

$$\frac{\alpha^2 - \omega^2}{2\alpha\omega} \sinh \alpha t_c \sin \omega t_o + \cosh \alpha t_c \cos \omega t_o + \cosh \alpha(t_c + t_o) = 0$$
$$. \quad . \quad . \quad . \quad (15)$$

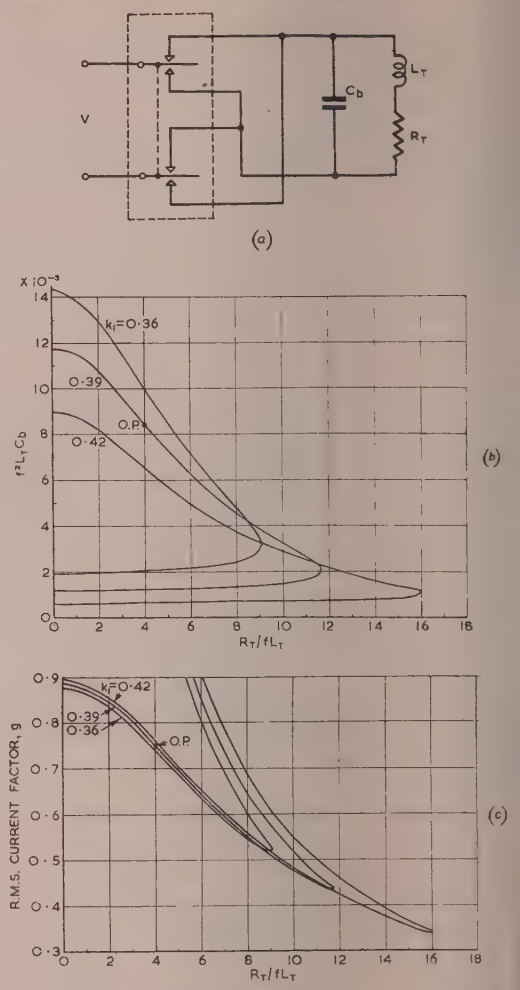


Fig. 9.—Equivalent circuit and characteristics, a.c. output, inductive load.

(*a*) Equivalent circuit referred to secondary winding.
(*b*) Relation between circuit parameters for ideal buffering.
(*c*) Output characteristics. Load current given by $I_{2R} = gV/2\pi fL_T$
O.P. Typical practical operating point.

where $\alpha = -R_T/2L_T$ and $\omega^2 = 1/L_TC_b - (R_T/2L_T)^2$. W t_c and t_o are written in terms of k_1 and f, the equation is to be in three variables k_1, $f^2L_TC_b$ and R_T/fL_T. In Fig. $f^2L_TC_b$ is plotted against R_T/fL_T for three values of k_1.

As in the case of no-load buffering, there are in general values of C_b for any particular value of R_T/fL_T, although the higher value is of practical importance, and for each v of k_1 there is a value of R_T/fL_T above which it is impossibl obtain the ideal waveform.

For a purely inductive load the circuit coincides with for no-load operation neglecting iron losses, so that Fig. may be employed, substituting L_T for L_n.

It should be noted that the capacitance C_b given by Fig. is that required across the secondary winding for a theoreti ideal waveform, and that in practice the value should be incre by about 50% in order to allow for wear of the contacts. M over, to determine the total capacitance it is necessary furth add that required to supply the transformer magnetizing cu (Section 3.2).

indicated in Section 11.2, an expression for the r.m.s. load nt in terms of the peak induced e.m.f. V is

$$I_{2R} = g\frac{V}{2\pi f L_T} \quad . \quad . \quad . \quad . \quad . \quad (16)$$

e g may be termed the "r.m.s. current factor" and is a ion of k_1 and R_T/fL_T. Values are plotted in Fig. 9(*c*).

e power dissipated in transformer and load is $R_T I_{2R}^2$, and is power is ultimately supplied by the battery at voltage V_1 lows that the mean input current (load component only) is by

$$I_1 = R_T I_{2R}^2 / V_1 \quad . \quad . \quad . \quad . \quad . \quad (17)$$

r.m.s. current in the primary winding, if centre-tapped, is ame in terms of I_{2R} as for a resistive load, namely

$$I_{w1} = nI_{2R}/\sqrt{2} \quad . \quad . \quad . \quad . \quad . \quad (18)$$

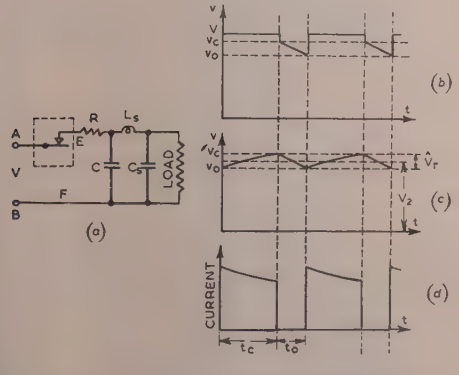
(5) D.C. OUTPUTS, MEAN RECTIFIED VOLTAGE

(5.1) Leakage Inductance Neglected

the majority of designs for power outputs of up to about atts the effect of transformer leakage inductance upon the ut voltage may be neglected; the output characteristics r these conditions will therefore be derived first, and the fying effect of inductance considered separately in on 5.2.

e equivalent circuit referred to the output is that shown in 10(*a*), and a constant voltage is assumed to be applied across erminals AB, although it is of course effective in supplying oad only during the portions of the cycle in which the ying contacts are closed. The applied voltage is the peak induced in the secondary winding at no load and is equal V_1 volts, or nV_1' if the voltage drop due to the no-load ary current may not be neglected (see Section 4.1.1). The ance R is the total circuit resistance referred to the output The following results apply also to circuits employing a ier instead of self-rectifying contacts, but in this case R des an allowance for the mean resistance of the rectifier, the closure-time ratio to be considered is that of the inter- r contacts (k_1) instead of the rectifying contacts (k_2).

e voltage and current waveforms are illustrated in 10(*b*)–10(*d*), and it is required to deduce an expression g the mean voltage V_2 across the reservoir capacitor C in s of the mean load current I_2. The problem is rather simpler



g. 10.—Equivalent circuit, d.c. output, neglecting leakage inductance of transformer.

(*a*) Equivalent circuit referred to output.
(*b*) Voltage across EF.
(*c*) Voltage across capacitor C.
(*d*) Current in resistor R.

than for the corresponding transformer-rectifier system operated from a sinusoidal supply, because the voltage is applied to the circuit for a fixed time (the closure-time of the contacts) and is constant during this interval. In the case of a sinusoidal supply,[6] however, the duration of each current pulse is dependent upon the circuit constants and the magnitude of the load current, because the rectifier begins to conduct when the applied voltage rises to equality with the capacitor voltage, and ceases to conduct at the instant when the current falls to zero.

In Fig. 10(*a*) choke L_s and capacitor C_s constitute the smoothing circuit. The percentage variation of voltage across C over a cycle is not large, and the percentage variation of current in L_s is even smaller, because the smoothing circuit and load present a high impedance to alternating current. Hence for the purpose of calculating the mean rectified voltage and the ripple voltage across the capacitor C it will be assumed that the current in L_s is constant throughout the cycle at the value I_2. The assumption is a usual one for capacitor-input rectifying systems, and moreover it is found in practice that for medium and light loads the calculation is sufficiently accurate even when there is no smoothing circuit beyond the capacitor C.

With the foregoing assumption it may be shown that the output characteristic is linear (see Section 11.3.1) so that the convertor behaves as a source of constant voltage (the open-circuit voltage) in series with an effective resistance R_e. Thus the internal power loss is $R_e I_2^2$, but since it is due to an r.m.s. current I_{2R} flowing in the circuit resistance R it follows that

$$R_e = R\left(\frac{I_{2R}}{I_2}\right)^2 \quad . \quad . \quad . \quad . \quad . \quad (19)$$

In words, the effective resistance of the convertor viewed from the output terminals is equal to the true resistance R of the circuit multiplied by the square of the form factor of the current flowing in R. The form factor is of course greater than unity, so that R_e is greater than R.

The open-circuit output voltage for full-wave rectification is nV_1, and hence the output characteristic has the equation

$$V_2 = nV_1 - R_e I_2 \quad . \quad . \quad . \quad . \quad . \quad (20)$$

It is convenient to write $R_e = sR \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$

where s may be termed the resistance multiplying factor, since it is the quantity by which the circuit resistance is multiplied to obtain the apparent or effective resistance. In Section 11.3.1 a general expression for s in terms of CR and the interval t_c and t_o is derived. For full-wave rectification it becomes

$$s = 2(1-k) + \frac{(1-2k)^2}{4fCR}\coth\left(\frac{k}{2fCR}\right) \quad . \quad . \quad (22)$$

which is a function of k and fCR. Values are plotted against $1/fCR$ in Fig. 11(*a*). Two special cases of interest are treated below.

If the reservoir capacitance is very large the voltage across it is almost constant over the whole cycle at the value V_2, so that the current in R consists of rectangular pulses of peak value $(V - V_2)/R$. The form factor of such a current is $1/\sqrt{(2k)}$, so that from eqn. (19) it follows that $s = 1/2k$. The same result may be obtained from eqn. (22) by finding the limit of s as $1/fCR$ approaches zero, remembering that $\coth x = 1/x$ for small values of x.

Reference to Fig. 11(*a*) shows that the curves of s are fairly flat for small values of $1/fCR$ (i.e. large capacitance) so that for a considerable range, say up to $1/fCR = 4$, the output is substantially independent of reservoir capacitance and vibrator frequency, and the effective resistance depends only upon R and k. Within this region a good approximation to the true value of s

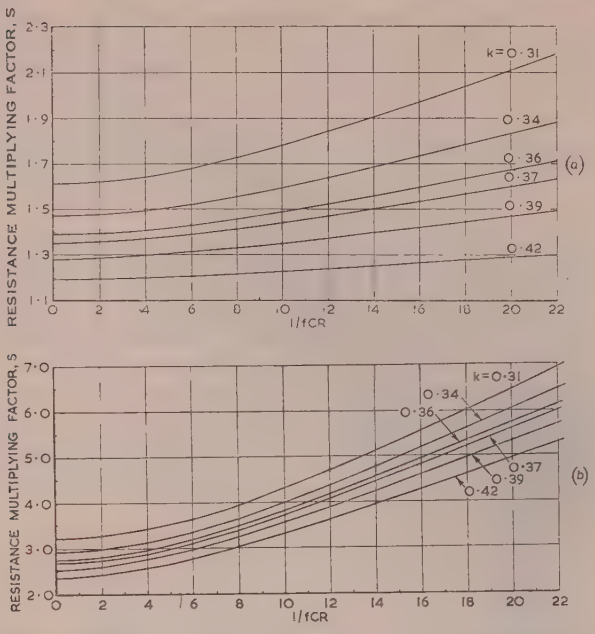


Fig. 11.—Effective resistance of convertor, d.c. output, neglecting leakage inductance $R_e = sR$.

(*a*) Full-wave.
(*b*) Half-wave.

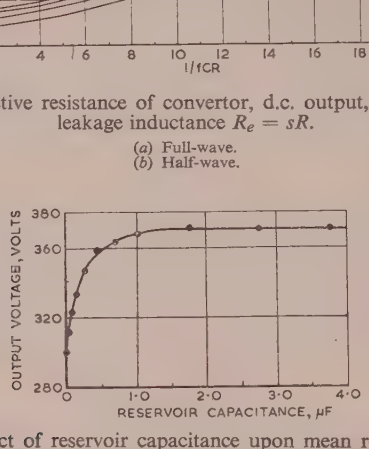


Fig. 12.—Effect of reservoir capacitance upon mean rectified output voltage (experimental results).

Full-wave rectification.
$f = 110$ c/s.
Constant load resistance = 4 630 ohms.

is given by $1/2k$. Fig. 12 gives test results obtained on a ty convertor when the voltage across a constant load resis was measured while the reservoir capacitance was varied will be noted that for capacitances above about 2 μF the vo is constant and therefore has the same value as with an in capacitance. Since in practice a capacitance of about would be provided in such a convertor, it would norr operate well within the region in which an assumption of in capacitance is justified. These conditions apply in the maj of designs employing full-wave rectification.

The other special case is when the circuit resistance R is small ($1/fCR$ large). Since $\coth x = 1$ for large values of x effective resistance becomes

$$R_e = sR = 2(1 - k)R + \frac{(1 - 2k)^2}{4fC} \quad . \quad .$$

so that even in the case of a circuit with negligible resist there would still be an effective resistance, having magn $(1 - 2k)^2/4fC$. This is because the reservoir capacitor sup the whole load and its voltage therefore falls, during the p that the vibrator contacts are open. This fall in voltage is portional to the load current, and the resulting fall in the r voltage over the cycle is also proportional to the load cur so that the net effect is that of the insertion of a resistance.

For half-wave rectification the output eqns. (20) and (21) a apply, but s is a different function of k and fCR. Value plotted in Fig. 11(*b*).

In a voltage-doubling circuit, each capacitor is part of a wave circuit, so that the output equation and the value of one capacitor is the same as for half-wave rectification. total output is twice that for each capacitor and is ther given by

$$V_2 = 2nV_1 - R_eI_2 \quad . \quad . \quad . \quad . \quad . \quad .$$

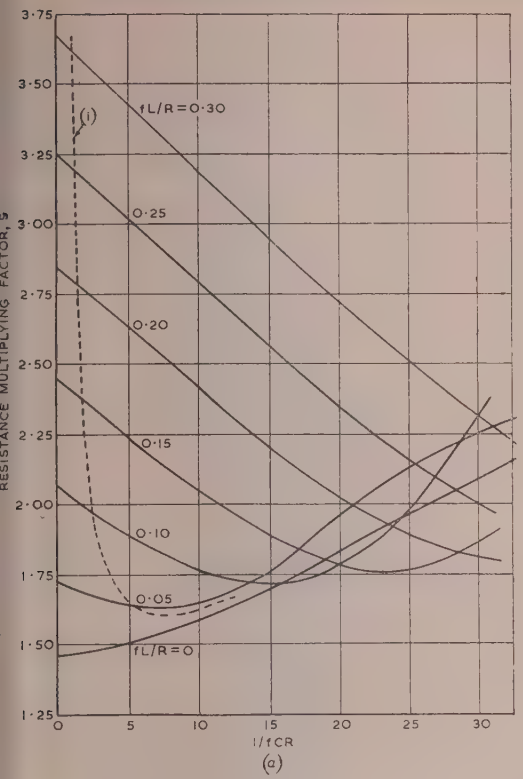
where now $R_e = 2sR$. The circuit resistance R is much sm than in a full-wave circuit of the same power output, so $1/fCR$ is much larger. Consequently, the effect of frequ and capacitance is more pronounced (the working point i longer on the flat part of the curves of s) and an assumptio infinite capacitance is not usually justified.

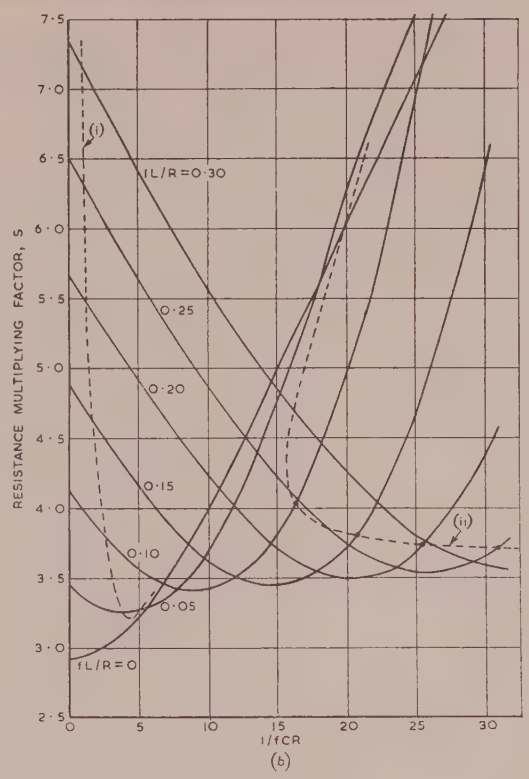
In all circuits, a decrease in frequency, reservoir capacit or closure-time ratio causes an increase in the effective resista although in different degrees as remarked above. The ou formulae are summarized in Table 1, the limiting cases

Table 1

MEAN RECTIFIED OUTPUT VOLTAGE NEGLECTING LEAKAGE INDUCTANCE

	Full-wave	Half-wave	Voltage-doubling
Output equation ..	$V_2 = nV_1 - R_eI_2$		$V_2 = 2nV_1 - R_eI_2$
Effective resistance, R_e	$R_e = sR$		$R_e = 2sR$
Resistance multiplying factor, s	$s = 2(1 - k) + \frac{(1 - 2k)^2}{4fCR} \coth\left(\frac{k}{2fCR}\right)$ Plotted in Fig. 11(*a*)	$s = 2 - k + \frac{(1 - k)^2}{2fCR} \coth\left(\frac{k}{2fCR}\right)$ Plotted in Fig. 11(*b*)	
s (C large)	$s = \frac{1}{2k}$	$s = \frac{1}{k}$	
R_e (R small)	$R_e = 2(1 - k)R + \frac{(1 - 2k)^2}{4fC}$	$R_e = (2 - k)R + \frac{(1 - k)^2}{2fC}$	$R_e = 2(2 - k)R + \frac{(1 - k$ fC





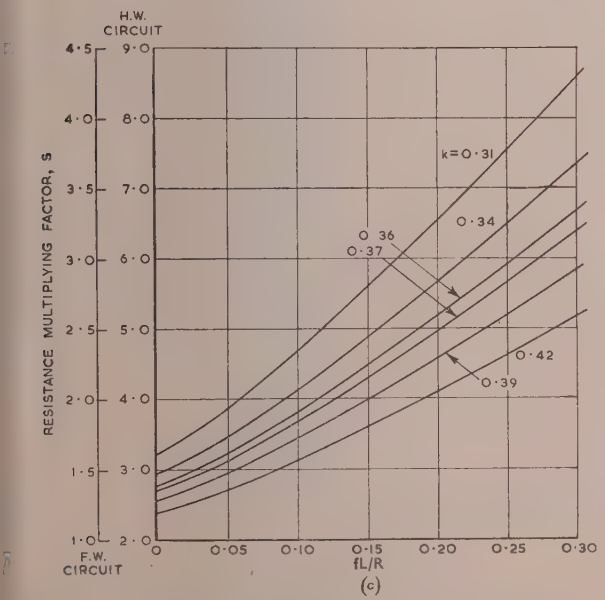


Fig. 13.—Effective resistance of convertor, d.c. output, considering leakage inductance $R_e = sR$.

(*a*) Full-wave, $k = 0{\cdot}34$.
(*b*) Half-wave, $k = 0{\cdot}34$.
(*c*) With infinite reservoir capacitance, full-wave and half-wave.
(i) Boundary of oscillatory region.
(ii) Curve along which contact unloading occurs.

included. The application of Table 1 in design involves
ıber of estimates, as is usual in such problems: it is necessary
timate the required transformer turns ratio, calculate the
ance of the resulting winding, calculate the output by
s of the formulae or curves, and compare this with the
d value. If it is not correct, a second estimate is then
, and so on.

nay be noted that V_2 is the mean voltage across the reservoir
itor and that the final output is less than this value by the
ge drop R_sI_2 in the winding of the smoothing choke L_s.

(5.2) Leakage Inductance considered

The equivalent circuit shown in Fig. 10(*a*) is now modified by the addition of the transformer leakage inductance referred to the secondary winding L, in series with the circuit resistance R.

It may be shown by the method given in Section 11.3.2 that the output characteristic is again linear and has the same equation as for the non-inductive case, except of course that the expressions for the multiplying factor s are different. The factor s is now a function of k, $1/fCR$ and fL/R; values for full-wave and half-wave rectification assuming $k = 0{\cdot}34$ are plotted in

Figs. 13(*a*) and 13(*b*). For a voltage-doubling circuit the curves relating to half-wave rectification are again employed. The figures include the curve having the equation $1/fCR = R/4fL$, which denotes the boundary between non-oscillatory and oscillatory charging of the reservoir capacitor; conditions are non-oscillatory in the region to the left of the curve.

It may be noted from Figs. 13(*a*) and 13(*b*) that each curve of s has a minimum value, and hence the output voltage is a maximum at some finite value of $1/fCR$, in contrast to the non-inductive case, in which the output attains a maximum for $1/fCR = 0$ (infinite capacitance). Moreover, for part of its length, a curve for high inductance may be below one of low inductance; over this range, therefore, high leakage inductance conduces to high output voltage. These results are due to a resonance effect between the leakage inductance and the reservoir capacitance, and are similar to those obtained for transformer-rectifier systems operating from a sinusoidal supply.[6]

Sometimes inductance is deliberately introduced to so shape the current pulses in the transformer that the contacts open at virtually zero current. This condition is fulfilled along the curve (ii) in Fig. 13(*b*). The subject is more fully discussed in Section 8.

For large values of reservoir capacitance the expression for the multiplying factor reduces to a simple function of k and fL/R (see Section 11.3.2). The function is plotted in Fig. 13(*c*), from which it may be noted that the value of s for half-wave rectification is twice that for full-wave rectification, other parameters being equal.

For a range of transformers designed for different frequencies but a given power, the value of fL/R increases with frequency. Thus the role played by leakage inductance in determining the regulation of convertors will be greater as vibrators of higher frequencies are employed. The value of fL/R also increases with size, i.e. with power rating. Moreover, it is higher for a voltage-doubling circuit than for a full-wave circuit. Typical figures for 110 c/s are 0·05 to 0·08 for full-wave rectification and approximately double these figures for the voltage-doubling circuit. Typical values of $1/fCR$ are 1 to 5 and 7 to 15 respectively for the two circuits.

(6) D.C. OUTPUTS, COPPER LOSS AND TRANSFORMER POWER RATING

(6.1) Currents and Copper Loss

The battery current corresponding to the rectified load current I_2 is given by

$$I_1 = nI_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

for full-wave and half-wave circuits

and by $I_1 = 2nI_2$ . . . . . ,

for voltage-doubling circuits. Once again, as for a.c. ou this is the load component, and to obtain the total cu flowing to the transformer the no-load component must be a

For all circuits the copper loss due to the load current is

$$P_k = R_e I_2^2 \quad . \quad . \quad . \quad . \quad . \quad .$$

Since this power is also equal to RI_{2R}^2, it follows that the current in the circuit resistance R for full- or half-rectification is given by

$$I_{2R} = \sqrt{(s)}I_2 \quad . \quad . \quad . \quad . \quad . \quad .$$

and for the voltage-doubling circuit by

$$I_{2R} = \sqrt{(2s)}I_2 \quad . \quad . \quad . \quad . \quad .$$

In order to determine the current density in the transfo windings it is necessary to derive expressions for the currents. With an untapped primary winding and rectification these currents are given by

$$I_{w2} = I_{2R} = \sqrt{(s)}I_2 \quad . \quad . \quad . \quad .$$

and $$I_{w1} = n\sqrt{(s)}I_2 = \sqrt{(s)}I_1 \quad . \quad . \quad . \quad .$$

On the other hand, if either winding is centre-tapped alternate current pulses flow in each half, and then the expressions are divided by $\sqrt{2}$.

The above results and the remaining expressions for a vo doubling circuit are summarized in Table 2. The half- circuit is not included, because owing to flux polarizatio magnetizing current assumes importance and invalidates s expressions for the current in the windings. The num value of s in Table 2 is read from the curves appropriate particular type of rectification, and the value of k is either k_2 according as an external rectifier or a self-rectifying vib is employed. The Table also includes the corresponding r for negligible leakage inductance combined with large rese capacitance, i.e. putting $s = 1/2k$ for full-wave rectifica and so on.

(6.2) Power Rating of Transformer

In Section 4.1.2 it was indicated that the power ratin transformers in different circuits may be compared by mea the coefficient λ in eqn. (13), and the numerical value of λ f a.c. output was computed. Similarly, with the aid of the ex sions for I_{w1} and I_{w2} given in Table 2, λ may be calculate the various types of rectifying circuit. The results are marized in Table 3. The particular case of negligible le

Table 2

CURRENTS FLOWING IN VIBRATORY CONVERTOR (RECTIFIED OUTPUT)

Rectifying circuit	I_{w1}/I_1				I_{w2}/I_2		I_1/I_2
	Centre-tapped primary		Double-pole change-over primary				
	(*a*)	(*b*)	(*a*)	(*b*)	(*a*)	(*b*)	
Full-wave (centre-tapped secondary) ..	$\frac{\sqrt{s}}{\sqrt{2}}$	$\frac{1}{2\sqrt{k}}$	$\sqrt{s}$	$\frac{1}{\sqrt{(2k)}}$	$\frac{\sqrt{s}}{\sqrt{2}}$	$\frac{1}{2\sqrt{k}}$	n
Full-wave (bridge)	$\frac{\sqrt{s}}{\sqrt{2}}$	$\frac{1}{2\sqrt{k}}$	$\sqrt{s}$	$\frac{1}{\sqrt{(2k)}}$	$\sqrt{s}$	$\frac{1}{\sqrt{(2k)}}$	n
Voltage-doubling..	$\frac{\sqrt{s}}{2}$	$\frac{1}{2\sqrt{k}}$	$\frac{\sqrt{s}}{\sqrt{2}}$	$\frac{1}{\sqrt{(2k)}}$	$\sqrt{(2s)}$	$\frac{\sqrt{2}}{\sqrt{k}}$	$2n$

I_{w1}, I_{w2} = r.m.s. currents in primary and secondary windings.
I_1, I_2 = battery current, load current (mean values).
(*a*) General. (*b*) With $L = 0$, $C = \infty$.

Table 3

POWER RATING OF VIBRATOR TRANSFORMERS, VALUE OF COEFFICIENT λ

Type of output	Centre-tapped primary			Double-pole change-over primary		
	(*a*)	(*b*)	(*c*)	(*a*)	(*b*)	(*c*)
ernating current	$\frac{8\sqrt{2}}{1+\sqrt{2}}\frac{\sqrt{k_1}}{1+2k_1}$	—	1·64	$4\sqrt{2}\frac{\sqrt{k_1}}{1+2k_1}$	—	1·98
l-wave rectification (centre-apped secondary)	$2\sqrt{2}\frac{1}{(1+2k_1)\sqrt{s}}$	$4\frac{\sqrt{k_2}}{1+2k_1}$	1·31	$\frac{8}{1+\sqrt{2}}\frac{1}{(1+2k_1)\sqrt{s}}$	$\frac{8\sqrt{2}}{1+\sqrt{2}}\frac{\sqrt{k_2}}{1+2k_1}$	1·54
l-wave rectification (bridge) ..	$\frac{8}{1+\sqrt{2}}\frac{1}{(1+2k_1)\sqrt{s}}$	$\frac{8\sqrt{2}}{1+\sqrt{2}}\frac{\sqrt{k_2}}{1+2k_1}$	1·54	$4\frac{1}{(1+2k_1)\sqrt{s}}$	$4\sqrt{2}\frac{\sqrt{k_2}}{1+2k_1}$	1·86
ltage-doubling rectification ..	$\frac{8\sqrt{2}}{1+\sqrt{2}}\frac{1}{(1+2k_1)\sqrt{s}}$	$\frac{8\sqrt{2}}{1+\sqrt{2}}\frac{\sqrt{k_2}}{1+2k_1}$	1·54	$4\sqrt{2}\frac{1}{(1+2k_1)\sqrt{s}}$	$4\sqrt{2}\frac{\sqrt{k_2}}{1+2k_1}$	1·86

(*a*) General, (*b*) $L = 0$, $C = \infty$. (*c*) $L = 0$, $C = \infty$, $k_1 = 0{\cdot}39$, $k_2 = 0{\cdot}34$.
For a sinusoidal input, and a.c. output, $\lambda = 2{\cdot}22$.

ctance and large reservoir capacitance is included, and these litions are then further particularized by putting $k_1 = 0{\cdot}39$ $k_2 = 0{\cdot}34$, thus obtaining a numerical value of λ. A self-fying vibrator is assumed except in the general case umns (*a*)]. The value of λ for an a.c. output is also included completeness. As is to be expected, the employment of re-tapped windings, where each half supplies power only ng alternate half-cycles, substantially reduces the permissible er rating of a given size of transformer.

(7) D.C. OUTPUTS, RIPPLE VOLTAGE

(7.1) Peak-to-Peak Ripple Voltage

uring an open interval of the contacts the voltage across the rvoir capacitor falls from v_c to v_0 [see Fig. 10(*c*)]. Except ighly inductive circuits such as are employed for contact ading (Section 8.3), v_c and v_0 are the maximum and minimum es of the instantaneous voltage during a closed interval also, at the peak-to-peak ripple voltage is $v_c - v_o$. If the current ing in the smoothing choke is again assumed to be sub-tially constant over the cycle at the value I_2, this ripple age is given by

$$\hat{V}_r = v_c - v_o = I_2 t_o / C \quad . \quad . \quad . \quad . \quad (32)$$

hen the appropriate values of t_o are substituted, the formula ull-wave rectification becomes

$$\hat{V}_r = \frac{1-2k}{2fC} I_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

that for half-wave rectification,

$$\hat{V}_r = \frac{1-k}{fC} I_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

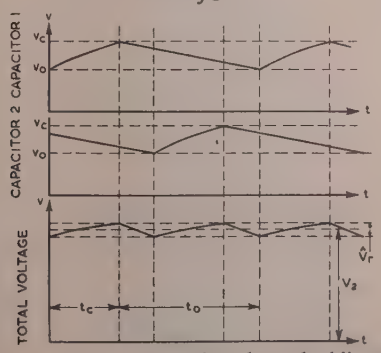


Fig. 14.—Voltage waveforms in voltage-doubling circuit.

In a voltage-doubling circuit the total output voltage rises while either vibrator contact is closed and falls while both are open (Fig. 14). In this latter interval, whose duration $(1 - 2k)/2f$ seconds is the same as in full-wave rectification, the load current is supplied by the two capacitors in series, so that the effective capacitance is $\frac{1}{2}C$ farads. The fall in voltage, and hence the peak-to-peak ripple, is thus given by

$$\hat{V}_r = \frac{1-2k}{fC} I_2 \quad . \quad . \quad . \quad . \quad . \quad (35)$$

Taking the ripple for full-wave rectification to be unity, that for a voltage-doubling circuit is seen to be 2, other factors being equal. On the same basis, and assuming $k = 0{\cdot}34$, the ripple for halfwave rectification is 4. It will be noted that the ripple voltage is directly proportional to load current, inversely proportional to vibrator frequency and reservoir capacitance, and substantially independent of the impedance and output voltage of the transformer.

(7.2) Harmonic Analysis of Ripple Voltage

The fundamental frequency of the ripple voltage is that of the vibrator for half-wave rectification, and is twice this value for full-wave or voltage-doubling circuits, provided that the contacts on either side of the reed have equal closure times.

In Section 11.4 the harmonic content of the ripple is analysed, neglecting the small modifying effect of transformer leakage inductance. In order to obtain the results in dimensionless form a ripple factor r_q is introduced, which expresses the r.m.s. value of the qth harmonic as a fraction of the peak-to-peak ripple voltage $\hat{V}_r$.

For a voltage-doubling circuit, the ripple harmonics across each capacitor are given by the formulae for half-wave rectification, but the net output ripple must be calculated by combining

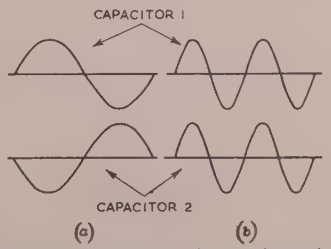


Fig. 15.—Harmonic analysis of ripple in voltage-doubling circuit.
(*a*) Fundamental. (*b*) Second harmonic.
The Figure illustrates the waveforms during one cycle of the vibrator and demonstrates that odd harmonics cancel, even harmonics add.

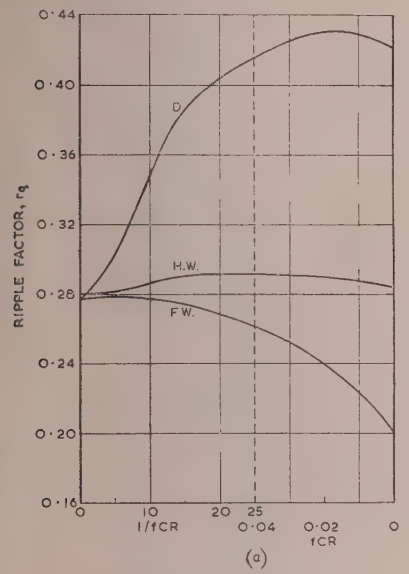
Table 4

RIPPLE VOLTAGE ACROSS RECTIFIED OUTPUT

R.M.S. value of qth harmonic $= r_q \hat{V}_r$

Rectifying circuit	Peak-to-peak ripple voltage $\hat{V}_r$	Fundamental frequency	Condition	Ripple factor, $r_q = \sqrt{(r_{sq}^2 + r_{cq}^2)}$	
				Sine component (r_{sq})	Cosine component (r_{cq})
Full-wave ..	$\frac{1-2k}{2fC} I_2$	$2f$	General	$\frac{\sqrt{2}}{2\pi q\sqrt{[1+(4\pi qfCR)^2]}}\left[\cos 2\pi qk + \frac{\sin 2\pi qk}{\pi q(1-2k)}\right]$	$\frac{2\sqrt{2fCR}}{\pi q\sqrt{[1+(4\pi qfCR)^2]}}\left(1+\frac{1-2k}{4fCR}\coth\frac{k}{2fCR}\right)\frac{\sin 2\pi qk}{1-2k}$
			fCR Large	Zero	$\frac{\sqrt{2}\sin 2\pi qk}{4\pi^2q^2k(1-2k)}$
			fCR Small	$\frac{\sqrt{2}}{2\pi q}\left[\cos 2\pi qk + \frac{\sin 2\pi qk}{\pi q(1-2k)}\right]$	$\frac{\sqrt{2}\sin 2\pi qk}{2\pi q}$
Half-wave ..	$\frac{1-k}{fC} I_2$	f	General	$\frac{\sqrt{2}}{2\pi q\sqrt{[1+(2\pi qfCR)^2]}}\left[\cos \pi qk + \frac{\sin \pi qk}{\pi q(1-k)}\right]$	$\frac{\sqrt{2fCR}}{\pi q\sqrt{[1+(2\pi qfCR)^2]}}\left(1+\frac{1-k}{2fCR}\coth\frac{k}{2fCR}\right)\frac{\sin \pi qk}{1-k}$
			fCR Large	Zero	$\frac{\sqrt{2}\sin \pi qk}{2\pi^2q^2k(1-k)}$
			fCR Small	$\frac{\sqrt{2}}{2\pi q}\left[\cos \pi qk + \frac{\sin \pi qk}{\pi q(1-k)}\right]$	$\frac{\sqrt{2}\sin \pi qk}{2\pi q}$
Voltage-doubling	$\frac{1-2k}{fC} I_2$	$2f$	General	$\frac{\sqrt{2}}{2\pi q\sqrt{[1+(4\pi qfCR)^2]}}\left[\cos 2\pi qk + \frac{\sin 2\pi qk}{2\pi q(1-k)}\right]\frac{1-k}{1-2k}$	$\frac{\sqrt{2fCR}}{\pi q\sqrt{[1+(4\pi qfCR)^2]}}\left(1+\frac{1-k}{2fCR}\coth\frac{k}{2fCR}\right)\frac{\sin 2\pi qk}{1-2k}$
			fCR Large	Zero	$\frac{\sqrt{2}\sin 2\pi qk}{4\pi^2q^2k(1-2k)}$
			fCR Small	$\frac{\sqrt{2}(1-k)}{2\pi q(1-2k)}\left[\cos 2\pi qk + \frac{\sin 2\pi qk}{2\pi q(1-k)}\right]$	$\frac{\sqrt{2}(1-k)\sin 2\pi qk}{2\pi q(1-2k)}$

ripples across the two capacitors, and these are not in phase. wever, it is shown in Fig. 14 that the waveform across one citor is the same as that across the other, but shifted by a period of the vibrator. As shown in Fig. 15, which



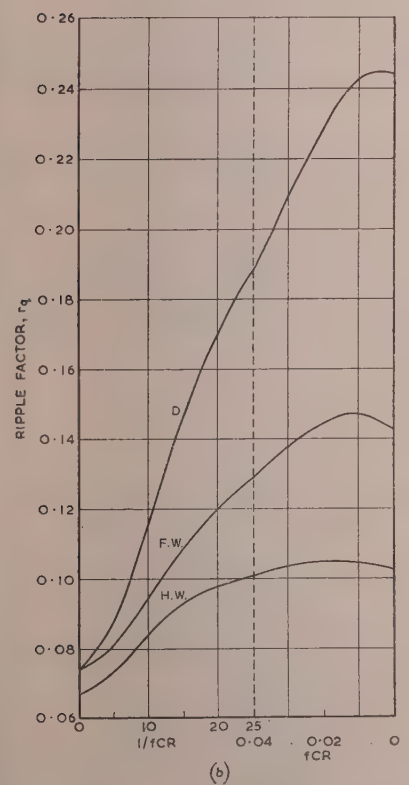


Fig. 16.—Magnitude of harmonics of ripple voltage.

(*a*) Fundamental.
(*b*) Second harmonic.
F.W. = Full-wave; H.W. = half-wave.
D = voltage-doubling circuit.
All curves are for $k = 0{\cdot}34$.
The r.m.s. value of qth harmonic $= r_q \hat{V}_r$.

employs the fundamental and the second harmonics as examples, this means that the odd harmonics are 180° out of phase so that they cancel each other, but the even harmonics are in phase so that they are additive. Thus the amplitude of the qth output harmonic is twice that of the $2q$th harmonic across one capacitor.

The formulae for peak ripple and ripple factor are given in Table 4 for the three circuits. The limiting cases $fCR \to \infty$ (large reservoir capacitance) and $fCR \to 0$ (small circuit resistance) are also included. In the former case both the charge and the discharge of the reservoir capacitor are substantially linear, and the ripple factors for the voltage-doubling and full-wave circuits are equal.

An examination of the formulae given in Table 4 indicates that the ripple factor for any circuit is a function of q, k and fCR only. In Fig. 16 the formulae have been plotted for the fundamental ($q = 1$) and the second harmonic ($q = 2$) for the three rectifying circuits, taking $k = 0{\cdot}34$. The higher harmonics are usually unimportant, especially when the reservoir capacitance is large.

Further smoothing of the ripple may be accomplished in the conventional manner by means of the choke L_s and the capacitor C_s [Fig. 10(*a*)], the harmonic of angular frequency $q\omega_f$ being attenuated by the factor

$$(q\omega_f)^2 L_s C_s - 1$$

(8) D.C. OUTPUTS WITH CONTACT UNLOADING FOR HIGH POWER OPERATION

(8.1) General

Dixey and Wilman[7] have investigated and described circuits in which the current pulses are so shaped that the contacts open virtually at no load. The main purposes are to achieve

(*a*) Greatly reduced transfer of contact material, enabling vibrator convertors to be operated at much higher powers than hitherto (e.g. 200–300 watts, compared with approximately 50 watts).
(*b*) Elimination of the high voltage peaks which normally occur owing to the interruption of the current flowing in the transformer leakage inductance. Hence the insulation of the vibrator and the transformer would be less highly stressed and radio interference would be reduced.

In this Section the series-resonant unloading circuit, as applied to convertors with rectified output, will be analysed.

The general effect of the transformer leakage inductance and reservoir capacitance on the rectified output voltage was discussed in Section 5.2. In a circuit employing a self-rectifying vibrator, if the reservoir capacitance is progressively decreased from a large value, all other parameters remaining constant, then, provided that the circuit inductance is sufficiently high, the waveform of the current flowing in the secondary winding changes in the manner indicated in Fig. 17, which shows only a

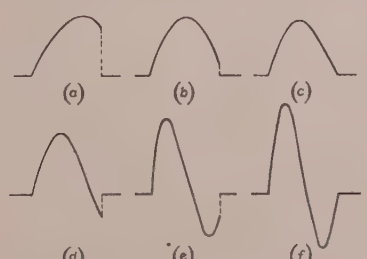


Fig. 17.—Waveforms of current flowing in rectifying contacts during one half-cycle of the vibrator.

(*a*)–(*f*) are for decreasing values of reservoir capacitance; contact unloading occurs (*c*) and (*f*).

half-cycle of vibrator operation. The primary current will have a similar waveform. It is seen that it is possible to choose the capacitance in relation to the inductance such that the current is zero at the instant the rectifying contacts open [(*c*) and (*f*) in Fig. 17]. In (*c*) the waveform has completed a half-cycle of the damped oscillation, in (*f*) a whole cycle. The latter is of course undesirable, as the reversed current subtracts from the mean rectified output current. The circuit inductance L required to achieve the desired waveform is usually greater than the transformer leakage inductance, so that a separate choke is usually included in series with the transformer secondary winding.

(8.2) Relationship between the Circuit Parameters

The necessary relationship between the circuit parameters for the current to take the waveforms (*c*) or (*f*) is given in Section 11.5.1 [eqn. (64)]. For a given type of rectification and contact closure-time ratio it may be expressed as a relation between f^2LC and fCR; this is plotted in Fig. 18 for a half-wave circuit with a range of values of k, and for a full-wave circuit with $k = 0{\cdot}34$.

It will be noted that the relationship depends only upon the parameters of the vibrator and the circuit; it is independent of the load current I_2, a useful feature when the load is variable. Mathematically, this independence is the consequence of the initial assumption that the current in the smoothing choke is substantially constant over the cycle. In other words, in order to render the unloading property independent of the load current, the load and smoothing circuit must have a high a.c. impedance compared to that of the reservoir capacitor. Thus the smoothing inductance L_s and the capacitance C should each be high, conditions which are desirable in any case in order to minimize the output ripple. If the impedance of the smoothing choke is not sufficiently high, it is found that at heavy overloads the contact unloading property is lost to some extent and the waveform approaches that of (*b*) in Fig. 17.

The curves in Fig. 18 show that in general there are two possible values of f^2LC for any value of fCR. The higher value corresponds to waveform (*c*), the lower to (*f*). Moreover for any value of k there is a maximum permissible value of fCR above which it is impossible to achieve the desired unloading property.

The practical application of the curves for half-wave rectification is in the voltage-doubling circuit, while the full-wave circuit usually employs a centre-tapped secondary winding, each half of which has twice the turns of the corresponding voltage-doubling winding. It may be shown that as a consequence the resistance of the full-wave circuit is about $5\frac{1}{2}$ times that of the doubling circuit. Since Fig. 18 indicates that the maximum permissible values of fCR in the two circuits have a ratio of about $1 : 2\frac{1}{2}$, it follows that the maximum permissible values of reservoir capacitance have a ratio of about 1 : 14. This low value of capacitance in a full-wave circuit necessitates high values of both series input inductance L and smoothing inductance L_s. Moreover, the size of the series choke will be further increased by the fact that it carries current having a unidirectional component. For these reasons a centre-tapped secondary winding is not usually employed and the voltage-doubling circuit is often favoured. An alternative would be a full-wave bridge circuit employing metal rectifiers or extra contacts on the vibrator, as this would eliminate some of the disadvantages of the centre-tapped circuit. It may also be mentioned that the use of a double-pole change-over input in place of a centre-tapped primary winding has the advantage of keeping R low and hence C high.

The choice of operating point on the curves of Fig. 18 will

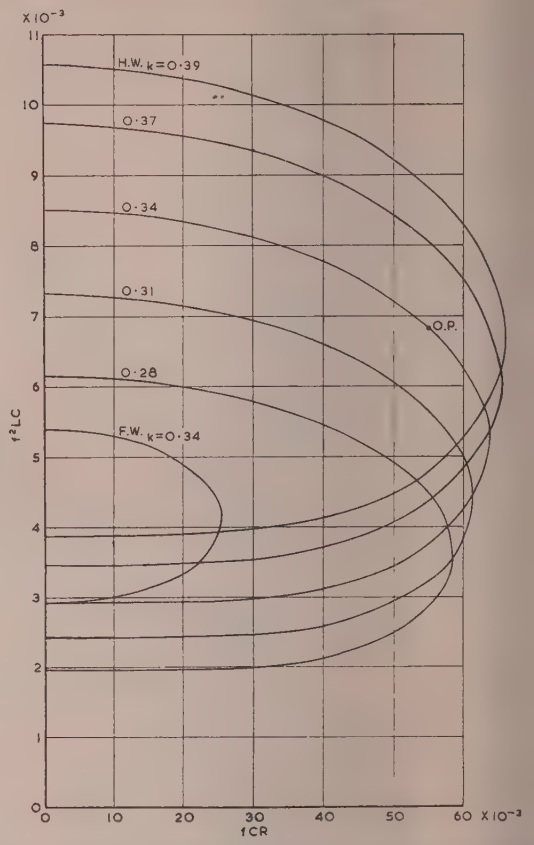


Fig. 18.—Relation between circuit parameters for contact unloa

All curves for half-wave (or voltage-doubling) circuit, except one marked full-wave.
O.P. Typical practical operating point.

now be considered. It will undoubtedly be on the upper of a curve, but it remains to be decided whether fCR sha large or small. Current waveforms for increasing value fCR are given in Fig. 19. When fCR is small the wavefo

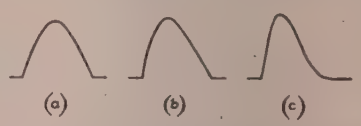


Fig. 19.—Current waveforms with contact unloading.

(*a*)–(*c*) are for increasing values of fCR.

nearly symmetrical, having the appearance of one half of a wave, and the rate of change of current at the moment the contact opens is high. As fCR increases the wave develops a trailing tail and its front becomes steeper. Fo limiting value of fCR (at the turning point of the curv Fig. 18) the rate of change, as well as the magnitude, o current is zero at the moment of opening. An approac this condition is desirable because it permits the timing o contacts to vary considerably during operating life without substantial current being interrupted. Thus it appears tha reservoir capacitance (and hence fCR) should be large, ticularly as this is also the condition which best accommo changes of load. Moreover, it requires a lower value of inductance, both because f^2LC is lower and because, even

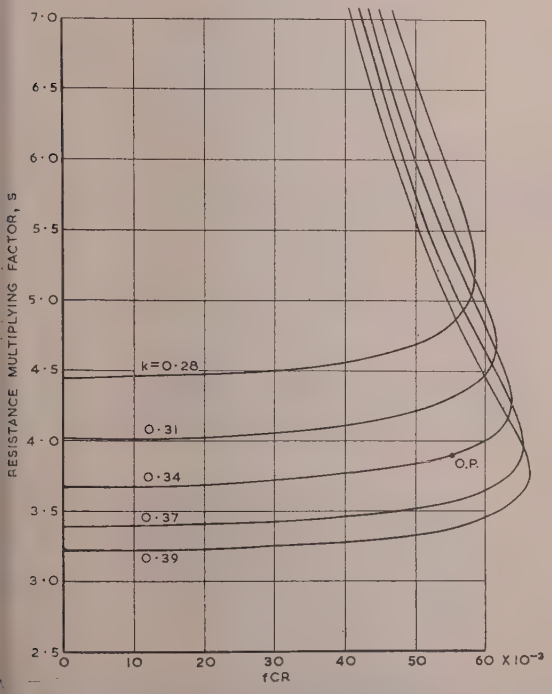
ven value of f^2LC, L is inversely proportional to C. There is e disadvantage, however. For small values of fCR the curves Fig. 18 are fairly flat, i.e. the current unloading property comparatively insensitive to changes in the circuit resistance. ow a high proportion of the circuit resistance is due to the rface resistance of the vibrator primary contacts reflected into output, and this quantity may be very variable. It follows at to avoid variations in the current waveform from this cause e reservoir capacitance should be small. One possible solution to make fCR large (about 0·9 of the permissible maximum) d to take measures to keep the contact resistance low and nstant, for example by filling the vibrator with a non-oxidizing s.

If a change in any of the parameters causes the operating int to move off its appropriate curve in Fig. 18, the current ll assume the waveform (*b*) or (*d*) (Fig. 17), depending upon ether the new operating point is above or below the curve.

(8.3) Mean Rectified Voltage and Ripple Voltage

The same linear output characteristics apply as in the more neral case dealt with in Section 5.2, but the expression for the istance multiplying factor s now assumes a simpler form qn. (67) of Section 11.5.2]. In the general case it was expressed a function of k, fCR and fL/R (Fig. 13), but now the relation-ip existing between the parameters allows it to be expressed à function of k and fCR only. Numerical values for a half-ve or a voltage-doubling circuit are plotted in Fig. 20. For voltage-doubling circuit the effective resistance R_e is again ven by $2sR$.

It may be noted that each curve in Fig. 20 is fairly flat until e turning point is approached, so that for a given transformer, d hence a given value of R, the choice of reservoir capaci-



g. 20.—Effective resistance of convertor, with contact unloading, half-wave, or voltage-doubling rectification.

Half-wave: $R_e = sR$.
Voltage-doubling: $R_e = 2sR$.
O.P. Typical practical operating point.

tance C is not important as far as the output voltage and the efficiency of the convertor are concerned. A curve of s against $1/fCR$ for half-wave rectification has been added to Fig. 13(*b*). A similar curve is not included in Fig. 13(*a*) (full-wave rectification) because the values of $1/fCR$ are outside the range covered in the diagram. Fig. 13(*b*) shows that for the larger values of fL/R the operating point for contact unloading occurs not far beyond the point of minimum s. In all cases the value of s with contact unloading is lower than it would be for zero inductance with the same values of C and R. Thus the addition of inductive impedance in order to achieve contact unloading actually increases the output for a given reservoir capacitance. The value of s is, however, always higher than it would be with zero inductance and very large capacitance ($1/fCR \to 0$).

The relative power rating of the transformer may be determined by substituting in the expressions of Table 3 [column (*a*)] the values of s given by eqn. (67).

It remains to determine the ripple voltage across the output before smoothing. Only the voltage-doubling circuit is dealt with, as this is the most common application. The peak-to-peak ripple $\hat{V}_r$ for the non-inductive or slightly inductive circuit has been derived in Section 7.1 [eqn. (35)]. The corresponding ripple $\hat{V}'_r$ for the contact unloading circuit is derived in Section 11.5.3, and the result is plotted in Fig. 21 in the form of the ratio $\hat{V}'_r/\hat{V}_r$ [eqn. (71)].

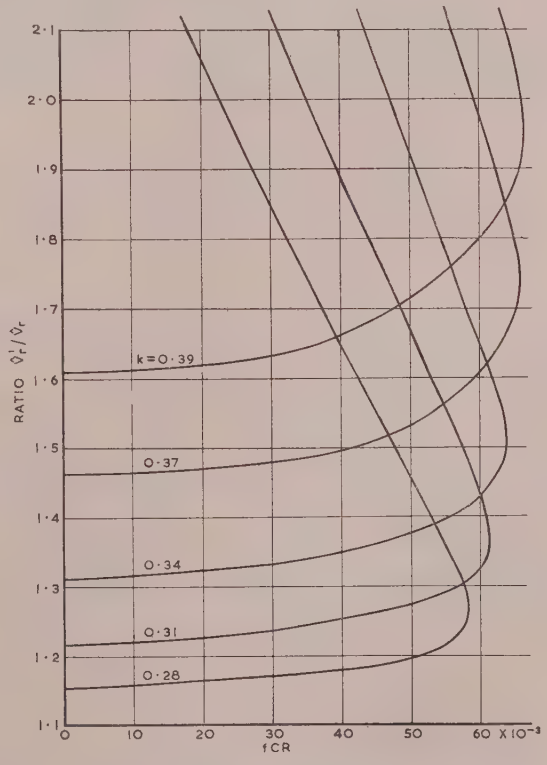


Fig. 21.—Peak-to-peak ripple voltage of voltage-doubling circuit, with contact unloading.

(9) ACKNOWLEDGMENTS

Acknowledgment is made to the Chief Scientist, the Ministry of Supply, and to the Controller of H.M. Stationery Office for permission to publish the paper.

(10) REFERENCES

(1) RADIO COMPONENTS STANDARDIZATION COMMITTEE: "Guide on Vibrators," RCG170, *Radio Components Standardization Committee* (Ministry of Supply, March, 1949).

(2) DISTIN, L. S.: "Modern Vibratory Power Convertors," *Post Office Electrical Engineers' Journal*, 1946, **39**, Part 2, p. 53.

(3) CONNELLY, F. C.: "Transformers" (Pitman, London, 1950), p. 326.

(4) MITCHELL, J. H.: "Recent Developments in Vibrators and Vibrator Power Packs," *Journal of the British Institution of Radio Engineers*, 1952, **12**, p. 431.

(5) ALLEN, A. L.: "Long-life Contacts for Unidirectional Currents of 1–20 Amperes," *Proceedings I.E.E.*, Paper No. 1506, July, 1953 (**100**, Part I, p. 158).

(6) TERMAN, F. E.: "Radio Engineering," 2nd Edition (McGraw-Hill, New York, 1937), p. 491.

(7) DIXEY, K. H., and WILMAN, C. V.: "Methods of increasing the Power Rating of Vibratory Convertors," *Proceedings I.E.E.*, Paper No. 1047R, March, 1951 (**98**, Part III, p. 105).

(11) APPENDICES

The method of analysis adopted is as follows. The voltage or current cycle may be divided into two distinct intervals, namely the closed interval of the vibrator contacts (t_c seconds) and the open interval (t_o seconds). The differential equations governing the waveforms in the two intervals are formulated and then solved, using the fact that each cycle is identical with the next once starting surges have died away; hence the initial and final conditions of the interval t_c are respectively equal to the final and initial conditions of the interval t_o.

(11.1) Analysis of Buffering Circuit

Consider the equivalent circuit of Fig. 3. During the interval t_c the instantaneous current in the inductance follows the relation

$$i = \frac{V_1}{L_n}t - i_o \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

where the initial condition $i = -i_o$ has been incorporated. During the interval t_o the instantaneous voltage across the circuit with initial condition $v = V_1$ is given by

$$v = V_1\varepsilon^{\alpha t}(G \sin \omega t + \cos \omega t) \quad . \quad . \quad . \quad (37)$$

where $\alpha = -1/2C_bR_n$, $\omega^2 = 1/L_nC_b - (1/2C_bR_n)^2$, and G is a constant to be determined. The expression for the instantaneous current i is of the same form.

The final current for the interval t_c is i_c, and the initial and final currents for the interval t_o are respectively i_c and i_o. Insertion of these conditions yields three simultaneous equations, the solutions of which are

$$\left.\begin{aligned} i_c &= \frac{V_1}{L_n}\frac{\omega t_c - \varepsilon^{\alpha t_o}\sin \omega t_o}{\omega - \varepsilon^{\alpha t_o}(\alpha \sin \omega t_o - \omega \cos \omega t_o)} \\ i_o &= \frac{V_1 t_c}{L_n} - i_c \\ G &= \frac{\alpha}{\omega} - \frac{i_c}{\omega C_b V_1} \end{aligned}\right\} \quad . \quad (38)$$

These expressions are valid for any value of buffer capacitance, and not merely for the optimum value.

For any buffer capacitance, the voltage across the circuit just before the contacts close is

$$V_1\varepsilon^{\alpha t_o}(G \sin \omega t_o + \cos \omega t_o) \quad . \quad . \quad (39)$$

For optimum buffering, however, this voltage is $-V_1$, so that

$$G = -\frac{1 + \varepsilon^{\alpha t_o}\cos \omega t_o}{\varepsilon^{\alpha t_o}\sin \omega t_o} \quad . \quad . \quad . \quad . \quad ($$

By substitution in eqn. (38) and elimination of i_c and i_o, expression relating the circuit parameters for optimum buffer may be shown to be

$$\frac{(\alpha^2 + \omega^2)t_c \sin \omega t_o}{2\omega} - \cos \omega t_o - \cosh \alpha t_o = 0 \quad .$$

For the non-oscillatory case put $\omega = j\beta$, and eqn. (5) th becomes

$$\frac{(\alpha^2 - \beta^2)t_c \sinh \beta t_o}{2\beta} - \cosh \beta t_o - \cosh \alpha t_o = 0 \quad . \quad ($$

(11.2) A.C. Outputs with Inductive Load

The equivalent circuit is shown in Fig. 9(*a*) and the wavefor are of the general shape shown in Fig. 2. During the inter t_c the load current is given by

$$i = \frac{V}{R_T} - \left(\frac{V}{R_T} + i_o\right)\varepsilon^{-R_T t/L_T} \quad . \quad . \quad . \quad ($$

and during the succeeding interval t_o the voltage across circuit follows the relation

$$v = V\varepsilon^{\alpha t}(G \sin \omega t + \cos \omega t) \quad . \quad . \quad . \quad ($$

where $\alpha = -R_T/2L_T$ and $\omega^2 = 1/L_TC_b - (R_T/2L_T)^2$. Th is a similar expression for the current. Insertion of the app priate boundary conditions yields the following relation betw the circuit parameters for an ideal trapezium-shaped wavefor

$$\frac{\alpha^2 - \omega^2}{2\alpha\omega}\sinh \alpha t_c \sin \omega t_o + \cosh \alpha t_c \cos \omega t_o + \cosh \alpha(t_c + t_o) =$$
$$. \quad . \quad . \quad . \quad . \quad ($$

For a purely inductive load ($R_T \to 0$) the equation reduces

$$\tfrac{1}{2}\omega t_c = \cot(\tfrac{1}{2}\omega t_o) \quad . \quad . \quad . \quad . \quad . \quad ($$

where $\omega^2 = 1/L_TC_b$.

The r.m.s. load current in the general case may be deduced follows. The quantity of electricity flowing from the source voltage V in a half-cycle is given by

$$\int_0^{t_c} i\,dt = \frac{Vt_c}{R_T} - \frac{L_T}{R_T}(i_c + i_o)$$

The power input is this quantity multiplied by $V/(t_c + t_o)$ a may be equated to $R_TI_{2R}^2$. Substitution of the values of i_c a i_o yields an expression for I_{2R} which is most convenien written

$$I_{2R} = g\frac{V}{2\pi fL_T} \quad . \quad . \quad . \quad . \quad . \quad . \quad ($$

where g is given by

$$g^2 = 4\pi^2\left(\frac{fL_T}{R_T}\right)^2\frac{t_c}{t_c + t_o}\left[1 - \frac{2\omega}{(\alpha^2 + \omega^2)t_c}\frac{\cosh \alpha t_o + \cos \omega t}{\sin \omega t_o}\right.$$
$$. \quad . \quad . \quad . \quad .$$

For a purely inductive load, eqn. (45) reduces to

$$g^2 = 2\pi^2[f^2L_TC_b + k_1^2(3 - 2k_1)/12] \quad . \quad .$$

where $f^2L_TC_b$ is the function of k_1 defined by eqn. (44) displayed graphically in Fig. 4(*b*).

(11.3) Mean Rectified Voltage

(1.3.1) **Leakage Inductance Neglected.**

The waveforms are illustrated in Fig. 10. Expressions for the pacitor voltage v during the intervals t_c and t_o respectively are

$$v = V - RI_2 - (V - RI_2 - v_o)\varepsilon^{-t/CR} \quad . \quad . \quad (47)$$

nd

$$v = v_c - I_2 t/C \quad . \quad . \quad . \quad . \quad . \quad (48)$$

he mean voltage across EF is

$$\frac{1}{t_c + t_o}\left[Vt_c + \frac{(v_c + v_o)t_o}{2}\right]$$

nd this expression may be equated to $V_2 + RI_2$, since I_2 is the ean current in R as well as in L_s. Substitution of the expres- ons for v_c and v_o derived by the insertion of the appropriate nal conditions in eqns. (47) and (48) yields the following pression for the mean rectified voltage:

$$V_2 = V - sRI_2 \quad . \quad . \quad . \quad . \quad . \quad (49)$$

here

$$s = 1 + \frac{t_o}{t_c + t_o}\left[1 + \frac{t_o}{2CR}\coth\left(\frac{t_c}{2CR}\right)\right] \quad . \quad . \quad (50)$$

hus the load characteristic is a straight line of slope sR.

For full-wave rectification $t_o = (1 - 2k)/2f$ and for half-wave ctification $t_o = (1 - k)/f$. For either circuit $t_c = k/f$.

(1.3.2) **Leakage Inductance considered.**

The instantaneous capacitor voltage during the interval t_c may ow be written

$$v = V - RI_2 + G\varepsilon^{m_1 t} + K\varepsilon^{m_2 t} \quad . \quad . \quad . \quad (51)$$

here

$$m = -\frac{R}{2L} \pm \sqrt{\left[\left(\frac{R}{2L}\right)^2 - \frac{1}{LC}\right]}$$

By integration of v over a cycle it may be shown that eqn. (49) gain applies but that s is now given by

$$= 1 - \frac{t_o}{t_c + t_o}\frac{t_o}{2CR} + \frac{L}{R(t_c + t_o)}\frac{(m_1 - m_2)(1 - E_1 - m_1 t_o)(1 - E_2 - m_2 t_o)}{m_1(1 - E_2) - m_2(1 - E_1)} \quad . \quad . \quad . \quad . \quad (52)$$

here $E_1 = \varepsilon^{m_1 t_c}$ and $E_2 = \varepsilon^{m_2 t_c}$.

The corresponding expression for oscillatory conditions is

$$= 1 - \frac{t_o}{t_c + t_o}\frac{t_o}{2CR} - \frac{\omega}{2\alpha(t_c + t_o)}\frac{(1 - \varepsilon^{\alpha t_c}\cos\omega t_c - \alpha t_o)^2 + (\varepsilon^{\alpha t_c}\sin\omega t_c + \omega t_o)^2}{\omega(1 - \varepsilon^{\alpha t_c}\cos\omega t_c) + \alpha\varepsilon^{\alpha t_c}\sin\omega t_c} \quad . \quad . \quad . \quad . \quad (53)$$

here $\alpha = -R/2L$ and $\omega^2 = 1/LC - (R/2L)^2$.

By allowing the value of C in eqn. (52) to increase indefinitely may be shown that for large reservoir capacitance the expres- on for s simplifies to

$$s = \frac{t_c + t_o}{t_c}\left[1 - \frac{L}{Rt_c}(1 - \varepsilon^{-Rt_c/L})\right]^{-1} \quad . \quad . \quad . \quad (54)$$

(11.4) Harmonic Analysis of Ripple Voltage

The voltage across EF in the equivalent circuit of Fig. 10(*a*) shown in Fig. 10(*b*). In Fig. 22 it is redrawn as a function the angle θ such that the period of one cycle is 2π instead of

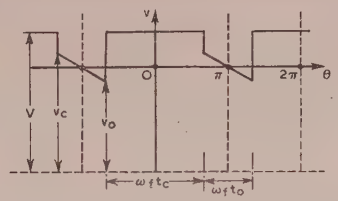


Fig. 22.—Voltage waveform across EF of Fig. 10 as function of θ.

$t_c + t_o$. The contacts are closed for an angle $\omega_f t_c$ and open for $\omega_f t_o$, where $\omega_f = 2\pi/(t_c + t_o)$.

The ripple voltage across EF may be expressed in the form of the Fourier series

$$v = \sum_{q-1}^{\infty} S_q \sin q\theta + \sum_{q-1}^{\infty} Q_q \cos q\theta \quad . \quad . \quad . \quad (55)$$

where the voltage and θ axes are placed as in Fig. 22 and the coefficients are defined by

$$\left.\begin{aligned} \pi S_q &= \int_0^{2\pi} v \sin q\theta d\theta \\ \pi Q_q &= \int_0^{2\pi} v \cos q\theta d\theta \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (56)$$

Employing the expressions for v_c and v_o derived during the calculation of the mean rectified voltage, it may be shown that

$$S_q = -\frac{v_c - v_o}{\pi\omega_f t_o}\int_{\frac{1}{2}\omega_f t_c}^{2\pi - \frac{1}{2}\omega_f t_c} (\theta - \pi)\sin q\theta d\theta = \frac{I_2 t_o}{\pi qC}\left(\cos\frac{q\omega_f t_c}{2} + \frac{2}{q\omega_f t_o}\sin\frac{q\omega_f t_c}{2}\right) \quad . \quad . \quad (57)$$

and

$$Q_q = \frac{2}{\pi}\int_0^{\frac{1}{2}\omega_f t_c}\left(V - \frac{v_c + v_o}{2}\right)\cos q\theta d\theta = \frac{2RI_2}{\pi q}\left(1 + \frac{t_o}{2CR}\coth\frac{t_c}{2CR}\right)\sin\frac{q\omega_f t_c}{2} \quad . \quad (58)$$

Between EF and the reservoir capacitor C the qth harmonic is attenuated in the ratio $\sqrt{[1 + (q\omega_f CR)^2]}$, so that the qth ripple harmonic (r.m.s. value) across C is the vector sum of the quadrature components,

$$\frac{S_q}{\sqrt{2}\sqrt{[1 + (q\omega_f CR)^2]}} \text{ and } \frac{Q_q}{\sqrt{2}\sqrt{[1 + (q\omega_f CR)^2]}}$$

Finally, each of these components may be expressed as a fraction of the peak-to-peak ripple voltage $\hat{V}_r$ by means of the ripple factors r_{sq} and r_{cq}. The total qth harmonic (r.m.s. value) is then given by $r_q\hat{V}_r$, where

$$r_q = \sqrt{(r_{sq}^2 + r_{cq}^2)} \quad . \quad . \quad . \quad . \quad . \quad (59)$$

It may be deduced from eqns. (57) and (58) that the component ripple factors are given by

$$r_{sq} = \frac{\sqrt{2}}{2\pi q\sqrt{[1 + (q\omega_f CR)^2]}}\left(\cos\frac{q\omega_f t_c}{2} + \frac{2}{q\omega_f t_o}\sin\frac{q\omega_f t_c}{2}\right) \quad . \quad . \quad . \quad . \quad (60)$$

and

$$r_{cq} = \frac{\sqrt{2CR}}{\pi q t_o\sqrt{[1 + (q\omega_f CR)^2]}}\left(1 + \frac{t_o}{2CR}\coth\frac{t_c}{2CR}\right)\sin\frac{q\omega_f t_c}{2} \quad . \quad . \quad . \quad . \quad (61)$$

(11.5) D.C. Outputs with Contact Unloading

(11.5.1) Equation for Contact Unloading.

Contact unloading can be achieved only if the circuit is oscillatory, i.e. if $(R/2L)^2 < 1/LC$. During the interval t_c the current flowing in the contacts and the voltage across the reservoir capacitor may be written

$$i = I_2 + I_2\varepsilon^{\alpha t}(G \sin \omega t - \cos \omega t) \quad . \quad . \quad . \quad (62)$$

$$v = V - RI_2 + LI_2\varepsilon^{\alpha t}$$
$$[G(\alpha \sin \omega t - \omega \cos \omega t) - \alpha \cos \omega t - \omega \sin \omega t] \quad . \quad (63)$$

where $\alpha = -R/2L$ and $\omega^2 = 1/LC - (R/2L)^2$. The discharge of the capacitor in the succeeding interval t_o follows the same relation as in the non-inductive case [eqn. (48)]. Insertion of all the boundary conditions establishes the following relation between the parameters for contact unloading,

$$\frac{(\alpha^2 + \omega^2)t_o \sin \omega t_c}{2\omega} - \cos \omega t_c + \cosh \alpha t_c = 0 \quad . \quad (64)$$

The expression for G is

$$G = -\frac{1 - \varepsilon^{\alpha t_c} \cos \omega t_c}{\varepsilon^{\alpha t_c} \sin \omega t_c} \quad . \quad . \quad . \quad . \quad (65)$$

In the limiting case of zero resistance ($\alpha \to 0$) eqn. (64) simplifies to

$$\tfrac{1}{2}\omega t_o = -\tan \tfrac{1}{2}\omega t_c \quad . \quad . \quad . \quad . \quad (66)$$

where $\omega^2 = 1/LC$.

(11.5.2) Mean Rectified Voltage.

The same output characteristic as before (Section 11.3.2) again applies, but the multiplying factor s is now given by

$$s = 1 + \frac{t_o}{2(t_c + t_o)}\left(1 - \frac{\omega \sinh \alpha t_c}{\alpha \sin \omega t_c}\right) \quad . \quad . \quad (67)$$

For zero resistance the corresponding formula is

$$s = \frac{3}{2} + \frac{t_c(\omega t_o)^2}{8(t_c + t_o)} \quad . \quad . \quad . \quad . \quad . \quad (68)$$

where $\omega^2 = 1/LC$.

(11.5.3) Peak Ripple, Voltage-Doubling Circuit.

When the total instantaneous voltage across the two capacito is a minimum or a maximum, one capacitor is charging at t same rate as the other is discharging. Thus each capacitor carr a current I_2 and the transformer at this instant is therefo supplying a current $2I_2$. Let the minimum and maximu voltages occur when ωt takes the values ψ_1 and ψ_2 respective time being measured from the beginning of the interval t_c. Th from eqn. (62) it follows that ψ_1 and ψ_2 are given by the solutio of the equation

$$\varepsilon^{\alpha\psi/\omega}(G \sin \psi - \cos \psi) = 1 \quad . \quad . \quad . \quad (6$$

where G is given by eqn. (65).

It may be shown that during the interval from ψ_1 to ψ_2 t total voltage rises by

$$\hat{V}_r' = \omega LI_2\left[\cot(\psi_1 - \gamma) + \tan\left(\psi_2 - \frac{\pi}{2} - \gamma\right)\right] - I_2(\psi_2 - \psi_1)/\alpha$$
$$. \quad . \quad . \quad . \quad . \quad (7$$

where $\tan \gamma = 1/G$, and this therefore is the peak-to-pe ripple. Alternatively it may be expressed as follows, as fraction of the ripple $\hat{V}_r$ for a non-inductive circuit,

$$\frac{\hat{V}_r'}{\hat{V}_r} = \frac{t_c}{t_o - t_c}\left\{\frac{\omega t_c}{(\alpha t_c)^2 + (\omega t_c)^2}\left[\cot(\psi_1 - \gamma) + \tan\left(\psi_2 - \frac{\pi}{2} - \gamma\right)\right] - \frac{\psi_2 - \psi_1}{\omega t_c}\right\} \quad . \quad (7$$

In the limiting case $R \to 0$ this expression simplifies to

$$\frac{\hat{V}_r'}{\hat{V}_r} = \frac{1 - 2\pi\sqrt{(f^2LC)}}{1 - 2k} \quad . \quad . \quad . \quad . \quad (7$$

21.372.54: 517.54

Monograph No. 110 M
Oct. 1954

A NOTE ON TIME SERIES AND THE USE OF JUMP FUNCTIONS IN APPROXIMATE ANALYSIS

By A. J. O. CRUICKSHANK, B.Sc., Ph.D., Associate Member.

The paper was first received 17th December, 1953, and in revised form 22nd March, 1954. It was published as an INSTITUTION MONOGRAPH *in October, 1954.)*

SUMMARY

The behaviour of filters having jump-function inputs is investigated. The output of the filter is approximated by a jump function, and the ratio between the Laplace transforms of the two jump functions defines the jump-transfer function of the filter. A serial operator for the filter can then be written down, and approximate analysis carried out using time series. A method of deriving a serial number for any time function having an analytical form is stated. Tables are given showing the jump-transfer function and serial operator for commonly occurring filters, and a comparison is made with the method due to Tustin of calculating approximate responses. A high degree of mathematical rigour has not been aimed at in the treatment.

LIST OF PRINCIPAL SYMBOLS

$f(t)$: Time function, having zero or non-zero initial value. $f(t) = 0, t < 0$.

$P(t)$: Rectangular pulse, height unity, width τ, starting at $t = 0$.

$\bar{P}(p)$: $\dfrac{1 - \varepsilon^{-\tau p}}{p}$, Laplace transform of $P(t)$.

$f_n = f(n\tau)$, ordinate of $f(t)$ at $t = n\tau$, $n = 0, 1, 2 \ldots$

$\lceil f(t)$: Jump function of $f(t)$; $\lceil f(t) = \sum_0^\infty f_n P(t - n\tau)$.

$u(t)$: Response of a filter to $P(t)$.

$H(t)$: Heaviside unit step function.

$A(t)$: Response of a filter to $H(t)$.

$z = \varepsilon^{\tau p}$.

p = Complex number.

(1) INTRODUCTION

The method, due to Tustin,[1] of analysing linear systems in terms of time series is particularly applicable when either the response to an arbitrary input, not representable in analytical form, is required, or when it is necessary to calculate the responses to any inputs, arbitrary or otherwise, of systems whose performance requires to be estimated from experimental tests. Such calculations are made by the expression of any time function by a series of numbers representing, in one method, the ordinates of that function at equal intervals, these being so chosen that a reasonable approximation of the time function is obtained; the effect of any system when subjected to an input so expressed is assessed by the serial operator for the particular system in question. The various methods of representing time functions by ordinates (Δ-units), step functions and rectangular pulse units are dealt with by Tustin. The serial operator for any particular system is shown to be obtained either from a known test input and the output it produces, or by the substitution $p = \dfrac{2}{\delta}\dfrac{[1, -1]}{[1, 1]}$ in the known transfer function of the system.

Brown[2] has described the application of finite-difference operators to linear systems and has shown in this connection that the operator $p = \dfrac{2}{\delta}\dfrac{(1 - E^{-1})}{(1 + E^{-1})}$, whose time series equivalent $\dfrac{2}{\delta}\dfrac{[1, -1]}{[1, 1]}$ was simply deduced by Tustin, is only one, and indeed the simplest, of a series of quotients of polynomials in E^{-1} which express the differential operator p to any desired degree of accuracy. More recently Barker[3] has drawn attention to the pulse-transfer function $W(z)$ of a filter, whose input and output are sequences of impulse functions, and has shown its application to sampling servo-systems. The material of this note relates to the transfer functions of filters whose inputs and outputs are jump functions. A jump-transfer function (j.t.f.) is derived, being very similar to the pulse-transfer function (p.t.f.) of the filter, and by means of the j.t.f. a serial operator for the filter is obtained. Likewise, the Laplace transform of a jump function may be used to derive, for a continuous time function of known analytical form, a serial number which is a quotient of two finite serial numbers, rather than an indefinite sequence of numbers representing the ordinates. Simple calculations are given to illustrate the method.

(2) THE JUMP-TRANSFER FUNCTION

(2.1) Basis

Consider the filter shown in Fig. 1, having the jump-function input $\lceil f(t)$.* $Y(p)$ is the transfer function of the filter and the

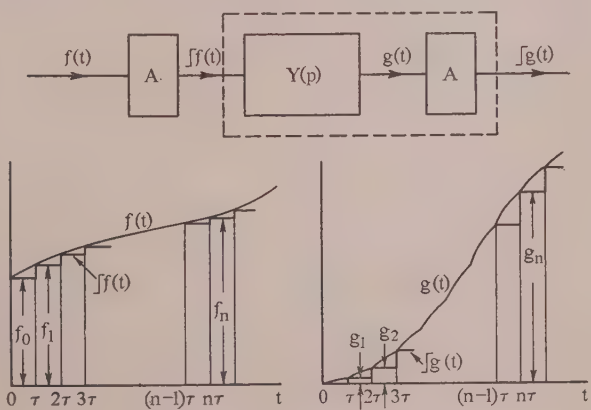


Fig. 1.—Filter with jump-function input.

boxes A are devices converting the continuous functions $f(t)$, $g(t)$ into the jump functions $\lceil f(t)$ and $\lceil g(t)$. The true output, $g(t)$, of the filter may not, however, have a continuous first derivative. It is desired to formulate the transfer function relating the Laplace transform of $\lceil g(t)$ to that of $\lceil f(t)$.

* The notation used for jump functions is that of Gardner and Barnes.[4]

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Cruickshank is at the University of St. Andrews.

Let $\bar{P}(p)$ be the Laplace transform of the rectangular pulse $P(t)$ of unit height and duration τ, starting at $t = 0$, i.e. $\bar{P}(p) = (1 - \varepsilon^{-\tau p})/p$. Then the transform of $\int f(t)$ is $\sum_0^\infty f_n \varepsilon^{-n\tau p}\bar{P}(p)$. Now let $u(t)$ be the response of $Y(p)$ to the unit pulse $P(t)$.

Then

$$g(t) = f_0 u(t) + f_1 u(t-\tau) + \ldots = \sum_0^\infty f_n u(t - n\tau)$$

and $\quad g(0) = f_0 u(0),\ g(\tau) = f_0 u(\tau) + f_1 u(0)$, etc.

If the ordinates of $u(t)$ at $t = 0, \tau, 2\tau, \ldots n\tau$ are denoted by $u_0, u_1, u_2, \ldots u_n$, then

$$g_n = f_0 u_n + f_1 u_{n-1} + f_2 u_{n-2} + \ldots + f_n u_0$$

Hence $\int g(t) = f_0 u_0 P(t) + (f_0 u_1 + f_1 u_0)P(t-\tau) + \ldots + (f_0 u_n + f_1 u_{n-1} + \ldots + f_n u_0)P(t - n\tau)$

and the transform of $\int g(t)$ is

$$\bar{P}(p)[f_0 u_0 + (f_0 u_1 + f_1 u_0)\varepsilon^{-\tau p} + \ldots + (f_0 u_n + f_1 u_{n-1} + \ldots + f_n u_0)\varepsilon^{-n\tau p} + \ldots] = \sum_0^\infty u_n \varepsilon^{-n\tau p} \sum_0^\infty f_n \varepsilon^{-n\tau p}\bar{P}(p)$$

If $z = \varepsilon^{\tau p}$, then the ratio [transform of $\int g(t)$]/[transform of $\int f(t)$] is given by $U(z) \fallingdotseq \sum_0^\infty u_n z^{-n}$.

$U(z)$ will be termed the jump-transfer function, denoted by j.t.f. for brevity. It is formally identical with the pulse-transfer function, see Barker,[3] given by $\sum_0^\infty w_n z^{-n}$. In the case of the j.t.f., the u_n are the ordinates of the response to a rectangular pulse of finite height unity and duration τ. In the case of the p.t.f., the w_n are the ordinates of the response to an impulse of finite area unity and infinitesimal duration.

(2.2) The Jump-Transfer Function of Common Filters

The jump-transfer function corresponding to the filter $Y(p) = 1/(p + \alpha)$ is deduced below for the purposes of illustration. It is necessary to find u_n, the sequence of ordinates representing the response to a unit rectangular pulse. This may be obtained by writing $u_0 = A_0$, $u_n = A_n - A_{n-1}$, $n = 1, 2, 3, \ldots$, where A_n is the sequence of ordinates representing the response to a unit step.

Therefore

$$\sum_0^\infty u_n z^{-n} = A_0 + (A_1 - A_0)z^{-1} + (A_2 - A_1)z^{-2} + \ldots + (A_n - A_{n-1})z^{-n} + \ldots = (1 - z^{-1}) \sum_0^\infty A_n z^{-n}$$

In this case $A(t) = (1 - \varepsilon^{-\alpha t})/\alpha$ and at $t = n\tau$, $A_n = (1 - d^n)/\alpha$ where $d = \varepsilon^{-\alpha\tau}$. Also $A_0 = 0$. The j.t.f. is therefore $(1 - d)z^{-1}/\alpha(1 - dz^{-1})$. This procedure may be repeated for other commonly occurring filters. Table 1 summarizes the results obtained. For all of the transfer functions given, $u(t)$ is zero at $t = 0$ and therefore the j.t.f. $\sum_0^\infty u_n z^{-n}$ reduces to $\sum_1^\infty u_n z^{-n}$. More complex transfer functions than those given may be split up into simpler ones by the use of partial fractions, and if desired, the overall j.t.f. may then be formed by summation.

(2.3) Filter with Finite Time-Delay

If a finite time-delay is added to a filter having the transfer function $Y(p)$, the resultant transfer function becomes $\varepsilon^{-\beta\tau p}Y(p)$, where $\beta\tau$ is the finite delay inserted. The same holds for th j.t.f., which becomes $z^{-\beta} \sum_0^\infty u_n z^{-n}$. The effect of the delay ca also be written in terms of the ordinates of the response to th unit pulse, the response now occurring $\beta\tau$ later in time. If β integral, then the response is zero up to $t = \beta\tau$, and for $t = m$ the ordinate of the response is $u_{n-\beta}$, where $u(t)$ is the respons of the original filter without finite time-delay. The j.t.f. of th filter with delay is therefore $\sum_{n=\beta}^\infty u_{n-\beta} z^{-n}$.

If β is a fraction of the interval τ an analogous result holds Let β be the fraction $(1 - m)$ of τ. Then at $t = \tau, 2\tau, 3\tau \ldots m$ the ordinate of the delayed response is $u(m\tau)$, $u(\tau + m\tau)$ $u(2\tau + m\tau)$ . . . $u(\overline{n-1}\tau + m\tau)$. The j.t.f. is therefor $\sum_{n=1}^\infty u_{n+m-1} z^{-n} = z^{-1} \sum_{n=0}^\infty u_{m+n} z^{-n}$.

(2.4) Laplace Transforms of Jump Functions

For either the exact analysis of systems in which jump function are the variables or, as will be shown later, the approximat analysis of systems having continuous variables, the Laplac transforms of jump functions may be used. Table 2 therefor shows the Laplace transforms of common jump-function These may be deduced in an elementary manner using th definition of the Laplace transform or by means of theorems o transform theory (see Gardner and Barnes[4]).

The inverse process of finding a jump function given it Laplace transform leads, as shown in Section 3, to the concept o serial numbers or time series, and also the concept of seria multiplication.

(3) TIME SERIES AND THE USE OF JUMP FUNCTIONS I APPROXIMATE ANALYSIS

(3.1) Inverse Laplace Transformation of Jump Functions: Seria Numbers

The inverse process of obtaining a jump function from it Laplace transform may be done in several ways. First, th transform may represent a known jump-function or be capabl of breaking up, by the process of partial fractions, into a numbe of recognizable jump-function transforms. This being a elementary method employed in normal transform theory, it wil not be discussed further.

A second approach is to expand the transform as a serie in z^{-1}. To take an example, consider the Laplace transform $\tau\bar{P}(p)\dfrac{z^{-1}}{(1 - z^{-1})^2}$. Expansion gives

$$\tau\bar{P}(p)z^{-1}[1 + 2z^{-1} + 3z^{-2} + \ldots + nz^{-(n-1)} + \ldots]$$

i.e. $\quad \tau[z^{-1}\bar{P}(p) + 2z^{-2}\bar{P}(p) + \ldots + nz^{-n}\bar{P}(p) + \ldots]$,

which is the Laplace transform of

$$\tau[P(t - \tau) + 2P(t - 2\tau) + \ldots + nP(t - n\tau) + \ldots]$$

$P(t)$ being the unit rectangular pulse of duration τ, starting a $t = 0$. This gives the required jump-function, in this case $\int$ The series may be obtained by simple division of the polynomial in z^{-1} in the numerator and denominator, if the latter is not th sum of a recognizable series.

It is also evident that the jump function $\int t$ may be describe by its ordinate heights at $t = 0, \tau, 2\tau \ldots$, etc. This gives, i this case, the sequence $\tau[0, 1, 2, \ldots]$, which is referred to a a serial number or time-series. In this sense, the power of z^-

Table 1

ELEMENTARY JUMP-TRANSFER FUNCTIONS

Transfer function	Ordinate of u(t) at $t = n\tau$, $n \geq 1$; $u_0 = 0$	Jump-transfer function
$Y(p)$	u_n	$\sum_{n=0}^{\infty} u_n z^{-n}$, $z = \varepsilon^{\tau p}$
$\frac{1}{p}$	τ	$\frac{\tau z^{-1}}{1 - z^{-1}}$
$\frac{1}{p^2}$	$\frac{\tau^2}{2!}(2n - 1)$	$\frac{\tau^2}{2!}\frac{(1 + z^{-1})z^{-1}}{(1 - z^{-1})^2}$
$\frac{1}{p^3}$	$\frac{\tau^3}{3!}(3n^2 - 3n + 1)$	$\frac{\tau^3}{3!}\frac{(1 + 4z^{-1} + z^{-2})z^{-1}}{(1 - z^{-1})^3}$
$\frac{1}{p^r}$	$\frac{\tau^r}{r!}[n^r - (n - 1)^r]$	$\frac{\tau^r}{r!}\frac{S_r(z^{-1})}{(1 - z^{-1})^r}$
$\frac{1}{p + \alpha}$	$\frac{d^{n-1}}{\alpha}(1 - d)$, $d = \varepsilon^{-\alpha\tau}$	$\frac{1 - d}{\alpha}\frac{z^{-1}}{(1 - dz^{-1})}$
$\frac{1}{(p + \alpha)^r}$	$d^{n-1}\sum_{q=0}^{r-1}\frac{\tau^q(n-1)^q}{\alpha^{r-q}q!} - d^n\sum_{q=0}^{r-1}\frac{\tau^q n^q}{\alpha^{r-q}q!}$	$\frac{z^{-1}}{\alpha^r} - \sum_{q=0}^{r-1}\frac{\tau^q(1 - z^{-1})}{\alpha^{r-q}q!}\frac{S_q(dz^{-1})}{(1 - dz^{-1})^{q+1}}$
$\frac{1}{p^2 + \omega^2}$	$\frac{1}{\omega^2}[\cos\omega(n-1)\tau - \cos\omega n\tau]$	$\frac{(1 - \cos\omega\tau)}{\omega^2}\frac{(1 + z^{-1})z^{-1}}{(1 - 2z^{-1}\cos\omega\tau + z^{-2})}$
$\frac{p}{p^2 + \omega^2}$	$\frac{1}{\omega}[\sin\omega n\tau - \sin\omega(n-1)\tau]$	$\frac{\sin\omega\tau}{\omega}\frac{(1 - z^{-1})z^{-1}}{(1 - 2z^{-1}\cos\omega\tau + z^{-2})}$
$\frac{1}{(p + \alpha)^2 + \omega^2}$	$\frac{1}{\omega r}\{d^{n-1}\sin[\omega(n-1)\tau + \phi] - d^n\sin[\omega n\tau + \phi]\}$ $d = \varepsilon^{-\alpha\tau}$, $\tan\phi = \frac{\omega}{\alpha}$, $r = (\alpha^2 + \omega^2)^{\frac{1}{2}}$	$\frac{Az^{-1} + Bz^{-2}}{\omega r^2(1 - 2dz^{-1}\cos\omega\tau + d^2z^{-2})}$ $A = \omega - \omega d\cos\omega\tau - \alpha d\sin\omega\tau$ $B = \omega d^2 - \omega d\cos\omega\tau + \alpha d\sin\omega\tau$
$\frac{p + \alpha}{(p + \alpha)^2 + \omega^2}$	$\frac{1}{r}\{d^{n-1}\cos[\omega(n-1)\tau + \phi] - d^n\cos[\omega n\tau + \phi]\}$	$\frac{Cz^{-1} + Dz^{-2}}{r^2(1 - 2dz^{-1}\cos\omega\tau + d^2z^{-2})}$ $C = \alpha - \alpha d\cos\omega\tau + \omega d\sin\omega\tau$ $D = \alpha d^2 - \alpha d\cos\omega\tau - \omega d\sin\omega\tau$
$\varepsilon^{-\beta\tau p}Y(p)$ β an integer	$u_{n-\beta}$	$z^{-\beta}\sum_{k=0}^{\infty}u_k z^{-k} = \sum_{n=\beta}^{\infty}u_{n-\beta}z^{-n}$
$\varepsilon^{-\beta\tau p}Y(p)$ β fractional equal to $(1 - m)$	u_{n+m-1}	$z^{-\beta}\sum_{k=0}^{\infty}u_k z^{-k} = \sum_{n=1}^{\infty}u_{n+m-1}z^{-n}$

* $S_q(dz^{-1}) = q!dz^{-1}$

$$\begin{vmatrix} 1 & 1 - dz^{-1} & 0 & \cdots & 0 \\ \frac{1}{2!} & 1 & 1 - dz^{-1} & \cdots & 0 \\ \frac{1}{3!} & \frac{1}{2!} & 1 & \cdots & 0 \\ \vdots & & & & \\ \frac{1}{q!} & \frac{1}{(q-1)!} & \frac{1}{(q-2)!} & \cdots & 1 \end{vmatrix}.$$

* See Reference 5.

Table 2

LAPLACE TRANSFORMS OF COMMON JUMP-FUNCTIONS: SERIAL NUMBERS

$$\int \mathrm{f}(t) = \sum_0^\infty f_n \mathrm{P}(t - n\tau),\ f_n = \mathrm{f}(n\tau)$$

Time function	Laplace transform	Jump function	Laplace transform	Serial number for time function
$\mathrm{P}(t)$	$\bar{\mathrm{P}}(p) = \dfrac{1 - \varepsilon^{-\tau p}}{p}$			$[1]$
$\mathrm{f}(t)\mathrm{H}(t)$	$\bar{\mathrm{f}}(p)$	$\int \mathrm{f}(t)$	$\bar{\mathrm{P}}(p)\sum_0^\infty f_n z^{-n},\ z = \varepsilon^{\tau p}$	$[f_0, f_1, f_2, \ldots]$
$A\mathrm{H}(t)$	$\dfrac{A}{p}$	$\int A\mathrm{H}(t)$	$A\bar{\mathrm{P}}(p)\dfrac{1}{1 - z^{-1}}$	$A\dfrac{[1]}{[1, -1]}$
$t\mathrm{H}(t)$	$\dfrac{1}{p^2}$	$\int t\mathrm{H}(t)$	$\tau\bar{\mathrm{P}}(p)\dfrac{z^{-1}}{(1 - z^{-1})^2}$	$\tau\dfrac{[0, 1]}{[1, -1]^2}$
$\dfrac{t^2}{2!}\mathrm{H}(t)$	$\dfrac{1}{p^3}$	$\int \dfrac{t^2}{2!}\mathrm{H}(t)$	$\dfrac{\tau^2}{2!}\bar{\mathrm{P}}(p)\dfrac{(1 + z^{-1})z^{-1}}{(1 - z^{-1})^3}$	$\dfrac{\tau^2}{2!}\dfrac{[1, 1][0, 1]}{[1, -1]^3}$
$\dfrac{t^3}{3!}\mathrm{H}(t)$	$\dfrac{1}{p^4}$	$\int \dfrac{t^3}{3!}\mathrm{H}(t)$	$\dfrac{\tau^3}{3!}\bar{\mathrm{P}}(p)\dfrac{(1 + 4z^{-1} + z^{-2})z^{-1}}{(1 - z^{-1})^4}$	$\dfrac{\tau^3}{3!}\dfrac{[1, 4, 1][0, 1]}{[1, -1]^4}$
$\varepsilon^{-\alpha t}\mathrm{H}(t)$	$\dfrac{1}{p + \alpha}$	$\int \varepsilon^{-\alpha t}\mathrm{H}(t)$	$\bar{\mathrm{P}}(p)\dfrac{1}{1 - dz^{-1}},\ d = \varepsilon^{-\alpha\tau}$	$\dfrac{[1]}{[1, -d]}$
$t\varepsilon^{-\alpha t}\mathrm{H}(t)$	$\dfrac{1}{(p + \alpha)^2}$	$\int t\varepsilon^{-\alpha t}\mathrm{H}(t)$	$\tau\bar{\mathrm{P}}(p)\dfrac{dz^{-1}}{(1 - dz^{-1})^2}$	$\tau\dfrac{[0, d]}{[1, -d]^2}$
$(\sin \omega t)\mathrm{H}(t)$	$\dfrac{\omega}{p^2 + \omega^2}$	$\int (\sin \omega t)\mathrm{H}(t)$	$\bar{\mathrm{P}}(p)\dfrac{z^{-1}\sin \omega\tau}{1 - 2z^{-1}\cos \omega\tau + z^{-2}}$	$\dfrac{[0, \sin \omega\tau]}{[1, -2\cos \omega\tau, 1]}$
$(\cos \omega t)\mathrm{H}(t)$	$\dfrac{p}{p^2 + \omega^2}$	$\int (\cos \omega t)\mathrm{H}(t)$	$\bar{\mathrm{P}}(p)\dfrac{(1 - z^{-1}\cos \omega\tau)}{1 - 2z^{-1}\cos \omega\tau + z^{-2}}$	$\dfrac{[1, -\cos \omega\tau]}{[1, -2\cos \omega\tau, 1]}$
$(\varepsilon^{-\alpha t}\sin \omega t)\mathrm{H}(t)$	$\dfrac{\omega}{(p + \alpha)^2 + \omega^2}$	$\int (\varepsilon^{-\alpha t}\sin \omega t)\mathrm{H}(t)$	$\bar{\mathrm{P}}(p)\dfrac{dz^{-1}\sin \omega\tau}{1 - 2dz^{-1}\cos \omega\tau + d^2z^{-2}}$ $d = \varepsilon^{-\alpha t}$	$\dfrac{[0, d\sin \omega\tau]}{[1, -2d\cos \omega\tau, d^2]}$
$(\varepsilon^{-\alpha t}\cos \omega t)\mathrm{H}(t)$	$\dfrac{p + \alpha}{(p + \alpha)^2 + \omega^2}$	$\int (\varepsilon^{-\alpha t}\cos \omega t)\mathrm{H}(t)$	$\bar{\mathrm{P}}(p)\dfrac{(1 - dz^{-1}\cos \omega\tau)}{1 - 2dz^{-1}\cos \omega\tau + d^2z^{-2}}$	$\dfrac{[1, -d\cos \omega\tau]}{[1, -2d\cos \omega\tau, d^2]}$

may simply be regarded as locating the position of the ordinate along the time axis. Hence the jump function is also adequately described by its original transform, the z^{-1} and $\bar{\mathrm{P}}(p)$ being omitted. That is, $\tau[0, 1, 2 \ldots] = \dfrac{\tau[0, 1]}{[1, -1]^2} = \dfrac{\tau[0, 1]}{[1, -2, 1]}$, the rule for evaluating $[1, -1]^2$ being the same as that for evaluating $(1 - z^{-1})^2$, namely multiplication of polynomials, as stated by Tustin.

This derivation of a serial number for a time function from the Laplace transform of the corresponding jump-function now makes possible the expression of the serial number in "closed form" rather than the straightforward ordinate sequence which, of course, may always be written down. Alternatively, it may be regarded as a summing process for ordinate sequences which are not otherwise recognizable. For example, the serial number (ordinates at $t = 0, \tau, 2\tau \ldots$) for $\varepsilon^{-\alpha t}\cos \omega t$ is $[1, \varepsilon^{-\alpha\tau}\cos \omega\tau, \varepsilon^{-2\alpha\tau}\cos 2\omega\tau, \ldots] = [1, d\cos \omega\tau, d^2\cos^2 \omega\tau \ldots]$ where $d = \varepsilon^{-\alpha\tau}$. This may be summed as it stands but it is also given by the Laplace transform of $\int \varepsilon^{-\alpha t}\cos \omega t$, omitting powers of z^{-1}, which are regarded as locating marks in the time axis, and leaving out $\bar{\mathrm{P}}(p)$. Thus the serial number for $\varepsilon^{-\alpha t}\cos \omega t$ is $\dfrac{[1, -d\cos \omega\tau]}{[1, -2d\cos \omega\tau, d^2]}$. Similarly the serial number for $\cos \omega t$ is $\dfrac{[1, -\cos \omega\tau]}{[1, -2\cos \omega\tau, 1]}$.

This also leads to the important effect which damping has on the serial number for, say, $\cos \omega t$. Referring to Table 2, if $\mathrm{C}(z^{-1})\mathrm{P}(\overline{p})$ is the transform of $\int \cos \omega t$, then $\mathrm{C}(dz^{-1})\bar{\mathrm{P}}(p)$ is the transform of $\int \varepsilon^{-\alpha t}\cos \omega t$; that is, dz^{-1} replaces z^{-1} exclusive of the part $\bar{\mathrm{P}}(p)$. The effect on the serial numbers, as can be seen from the examples, is to change the ordinates from f_n to $d^n f_n$, $n \geqslant 1$, whether the numbers are infinite sequences or quotients of finite sequences.

A third method of evaluating the transforms of a jump function is to consider it as the product of two simpler transforms representing known jump-functions. Use is then made of the theorem regarding the transform of the convolution of two time functions. For a simple example consider the transform $\tau\bar{\mathrm{P}}(p)\dfrac{z^{-1}}{(1 - z^{-1})^2}$. Let $\bar{x}_1(p) = \dfrac{\bar{\mathrm{P}}(p)}{1 - z^{-1}}$, which represents the transform of the unit step. Let $\bar{x}_2(p) = \dfrac{z^{-1}}{(1 - z^{-1})} = z^{-1} + z^{-2} + z^{-3} + \ldots$ Then $\bar{x}_2(p)$ is the transform of $\delta(t - \tau) + \delta(t - 2\tau) + \delta(t - 3\tau) \ldots$, $\delta(t)$ being the unit impulse, whose transform is unity. According to the convolution theorem,* therefore, $\bar{x}_1(p)\bar{x}_2(p)$ is the transform of $\int_0^t x_2(\lambda)x_1(t - \lambda)d\lambda$. The result of the integration is

* See Reference 4, p. 228.

$x_1(t-\tau) + x_1(t-2\tau) + \ldots = y(t)$, say, which gives the sequence $[y_0, y_1, y_2 \ldots] = [0, 1, 2, 3 \ldots]$. It may be verified that this is the sequence resulting from multiplication of $(z^{-1} + z^{-2} + z^{-3} + \ldots)$ by $(1 + z^{-1} + z^{-2} + \ldots)$, which represents the unit step. Serial multiplication of sequences can therefore be looked upon as a numerical convolution process.

(3.2) Use of Jump-Transfer Function in Approximate Analysis

There are two obvious methods whereby a jump function may be used to approximate a continuous function. The first approximation is one in which the heights of the rectangular pulses are the averages of the two adjacent ordinates of the curve marking the beginning and end of the rectangular pulse in question. The second method is that in which the heights of the rectangular pulses are given by the ordinates of the curve at the mid-points of the pulses. In the paper, more consideration is given to the first method, which in fact amounts to calculation in terms of the ordinates at $t = 0, \tau, 2\tau \ldots$, etc. It is desired, therefore, in the first place, to derive from the j.t.f. a serial operator which is the ratio of the serial numbers of the output and input when these are the simple ordinate sequences.

Consider Fig. 2 in which the input f(t) to a filter is approxi-

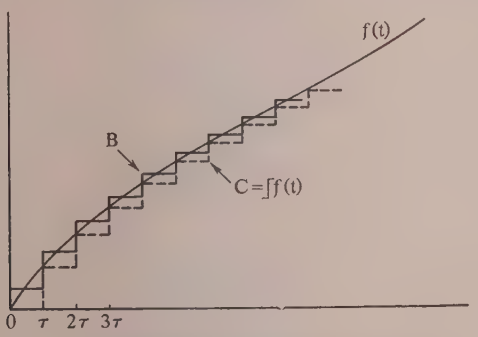


Fig. 2.—Approximation of curve by jump function.

mated by the jump function B, having the Laplace transform $\bar{B}(z)$. If $U(z)$ is the j.t.f. for the filter, the jump function whose pulse heights are equal to the ordinates of the output at $t = 0$, $\tau, 2\tau, \ldots$, etc., has the transform $U(z)\bar{B}(z) = \bar{D}(z)$, say. But $\bar{B}(z) = \frac{1}{2}(1 + z)\bar{C}(z)$ where $\bar{C}(z)$ is the transform of the jump function whose pulse heights are equal to the ordinates of f(t) at $t = 0, \tau, 2\tau, \ldots$, etc. The serial operator for the filter, equal to the ratio (serial number of output)/(serial number of input), is therefore given by $\bar{D}(z)/\bar{C}(z) = \frac{1}{2}(1 + z)U(z)$, omitting all powers of z^{-1}. For example, consider the transfer function $\alpha/(p + \alpha)$ which has the j.t.f. $(1 - d)z^{-1}/(1 - dz^{-1})$. The serial operator is obtained from $\frac{1}{2}(1 + z)U(z)$ when the powers of z^{-1} are regarded as locating points on the time axis and are now omitted. This gives the serial operator $\frac{1}{2}(1 - d)[1, 1]/[1, -d]$ the value of $d = \varepsilon^{-\alpha\tau}$ depending on the value of τ adopted. If τ equals one-fifth of the time-constant $1/\alpha$, then $d = 0{\cdot}8187$ and the serial operator is $0{\cdot}0906[1, 1]/[1, -0{\cdot}8187]$. If the input is a ramp function, say, then its serial number is $\tau[0, 1, 2, 3, \ldots]$ and the output is $0{\cdot}0906\tau[0, 1, 2, 3 \ldots] \times [1, 1]/[1, -0{\cdot}8187]$. Evaluating gives $\tau[0, 0{\cdot}091, 0{\cdot}347, 0{\cdot}737, 1{\cdot}24, 1{\cdot}83, 2{\cdot}5, 3{\cdot}22, 4{\cdot}0 \ldots]$ compared with the exact solution $\tau[0, 0{\cdot}093, 0{\cdot}351, 0{\cdot}744, 1{\cdot}246, 1{\cdot}84, 2{\cdot}51, \ldots]$.

To summarize, therefore, the procedure of approximate analysis, the steps are: (*a*) Derive the serial operator for the particular transfer function; this is obtained from $\frac{1}{2}(1 + z)$ times the selected j.t.f. from Table 1, bearing in mind that more complex transfer functions have j.t.f.'s which may be built up from the simple ones listed.

(*b*) Multiply the serial operator by the serial number representing the input; this may take the form either of the ordinate heights at $0, \tau, 2\tau, \ldots$, etc., or of a quotient of two finite serial numbers, as given by Table 2.

(*c*) Evaluate the serial product so obtained by division, multiplication or otherwise

The above procedure will hold in calculating the responses only to inputs which have a zero initial value. That this should be so is evident if it is remembered that the transform of the jump-function approximation* to a function f(t) is

$$\tfrac{1}{2}\{[1 + z][\text{transform of } \int \mathrm{f}(t)] - z\mathrm{f}(0)\bar{\mathrm{P}}(p)\}$$

Thus only when f(0) = 0 may the factor $\frac{1}{2}(1 + z)$ be taken along with the j.t.f. in deriving a serial operator which is to be the quotient of the ordinate sequences of the output and input. The lack of generality does not, however, constitute any difficulty, for inputs with non-zero initial values may always be split up into a step function and an input with a zero initial value. In the case of a step function the question of approximate analysis does not arise, for the step function is the exact jump-function $A[1, 1, 1, \ldots]$, A being the height of the step. The serial operator for a step-function input is therefore the j.t.f. for the filter, omitting the powers of z^{-1} as before. For example, the output of the filter $\alpha/(p + \alpha)$, when subjected to the input $(1 + t)H(t)$, and with $\tau = 1/5\alpha$, is

$$\frac{\tau(1-d)}{2}\frac{[1,1]}{[1,-d]}[0,1,2,3\ldots] + (1-d)\frac{[0,1]}{[1,-d]}[1,1,1,\ldots]$$
$$= \tau[0{\cdot}091, 0{\cdot}347, 0{\cdot}737, 1{\cdot}24 \ldots]$$
$$+ [0, 0{\cdot}181, 0{\cdot}329, 0{\cdot}451, 0{\cdot}550 \ldots]$$

It is not proposed to deal in detail with the second method of approximating a continuous function by means of a jump function, that is, where the pulse heights are equal to the ordinates of the function at $t = \tau/2, 3\tau/2, \ldots (n + \frac{1}{2})\tau \ldots$ The procedure is very similar to that of Section 2.1 and the results are formally the same. Thus if a jump function $\sum_{n=0}^{\infty} F_n \mathrm{P}(t - n\tau)$ is applied to a filter and the output G(t) is approximated by the jump function $\sum_{n=0}^{\infty} G_n \mathrm{P}(t - n\tau)$, where in this case $G_n = \mathrm{G}[(n + \frac{1}{2})\tau]$, then the j.t.f. is still $\sum_{0}^{\infty} u_n z^{-n}$, but here $u_n = \mathrm{u}[(n + \frac{1}{2})\tau]$, i.e. the ordinate at $t = (n + \frac{1}{2})\tau$ of the response of the filter to a unit pulse of width τ.

The application to approximate analysis follows simply, for if F_n is now equal to $\mathrm{F}[(n + \frac{1}{2})\tau]$, i.e. the second type of approximation to an actual input F(t), the approximation is consistent for both input and output, and the serial operator is the j.t.f. omitting the powers of z^{-1} as before. This applies for any input. The j.t.f. for the filter is of course slightly different from that previously used, since u_n is now $\mathrm{u}[(n + \frac{1}{2})\tau]$ instead of $\mathrm{u}(n\tau)$. For instance the j.t.f. for the filter $1/(p + \alpha)$ is now

$$\frac{1}{\alpha} - \frac{d^{\frac{1}{2}}(1 - z^{-1})}{\alpha(1 - dz^{-1})}$$

where $d = \varepsilon^{-\alpha\tau}$. This method introduces rather more complex j.t.f.'s than the first approach and it is not quite so convenient to work in terms of the ordinates at $t = (n + \frac{1}{2})\tau$ as to use those at $t = n\tau$. The process does hold, however, for all inputs, whether their initial values are zero or not.

* That is, pulse heights of jump functions are averages of ordinates on either side of pulse.

(3.3) Comparison with Method of Substituting $p = \frac{2}{\tau}\frac{[1,-1]}{[1,1]}$ in Transfer Function

The method used by Tustin to calculate approximate responses is to substitute $p = \frac{2}{\tau}\frac{[1,-1]}{[1,1]}$ in the transfer function of the system. This rule enables a serial operator to be obtained for a general form of the transfer function such as the quotient of two polynomials in p. Clearly this procedure is much simpler than breaking the transfer function up into simpler ones by a method of partial fractions and then obtaining a serial operator for each.

The comparison may also be made using the transfer functio $\alpha/(p+\alpha)$, which gives the serial operator $(1-d)[1,1]/2[1,-d]$ where $d = \varepsilon^{-\alpha\tau}$. Expanding in terms of $\alpha\tau$, using powers u to $\alpha^3\tau^3$, this may be written

$$\frac{1}{2}\left(\alpha\tau - \frac{\alpha^2\tau^2}{2} + \frac{\alpha^3\tau^3}{6}\right)\frac{[1,1]}{\left[1, -\left(1 - \alpha\tau + \frac{\alpha^2\tau^2}{2} - \frac{\alpha^3\tau^3}{6}\right)\right]}$$

If the substitution $p = \frac{2}{\tau}\frac{[1,-1]}{[1,1]}$ is made in $Y(p)$ and a simila

Table 3

COMPARISON OF SERIAL OPERATORS

$Y(p)$	Serial operator using $p = \frac{2}{\tau}\frac{[1,-1]}{[1,1]}$	Serial operator* using $\frac{(1+z)}{2}$(jump-transfer function)
$\frac{1}{p}$	$\frac{\tau}{2}\frac{[1,1]}{[1,-1]}$	$\frac{\tau}{2}\frac{[1,1]}{[1,-1]}$
$\frac{1}{p^2}$	$\frac{\tau^2}{4}\frac{[1,1]^2}{[1,-1]^2}$	$\frac{\tau^2}{4}\frac{[1,1]^2}{[1,-1]^2}$
$\frac{1}{p^3}$	$\frac{\tau^3}{8}\frac{[1,3,3,1]}{[1,-1]^3}$	$\frac{\tau^3}{12}\frac{[1,5,5,1]}{[1,-1]^3}$
$\frac{1}{p+\alpha}$	$\frac{\tau}{2+\alpha\tau}\frac{[1,1]}{[1,-b]}$, $b = \frac{(2-\alpha\tau)}{(2+\alpha\tau)}$	$\frac{(1-d)}{\alpha}\frac{[1,1]}{[1,-d]}$, $d = \varepsilon^{-\alpha\tau}$
$\frac{1}{p^2+\omega^2}$	$\frac{\tau^2}{4+\omega^2\tau^2}\frac{[1,1]^2}{[1,-2c,1]}$ $c = (4-\omega^2\tau^2)/(4+\omega^2\tau^2)$	$\frac{(1-\cos\omega\tau)}{2\omega^2}\frac{[1,1]^2}{[1,-2\cos\omega\tau,1]}$
$\frac{p}{p^2+\omega^2}$	$\frac{2\tau}{4+\omega^2\tau^2}\frac{[1,0,-1]}{[1,-2c,1]}$	$\frac{\sin\omega\tau}{2\omega}\frac{[1,0,-1]}{[1,-2\cos\omega\tau,1]}$
$\frac{1}{(p+\alpha)^2+\omega^2}$	$\frac{\tau^2}{g}\frac{[1,2,1]}{\left[1,-2f/g,1-\frac{8\alpha\tau}{g}\right]}$ $f = 4 - r^2\tau^2$ $g = 4 + 4\alpha\tau + r^2\tau^2$ $r^2 = \alpha^2 + \omega^2$	$\frac{A}{2\omega r^2}\frac{[1,1+B/A,B/A]}{[1,-2d\cos\omega\tau,d^2]}$ $A = \omega - \omega d\cos\omega\tau - \alpha d\sin\omega\tau$ $B = \omega d^2 - \omega d\cos\omega\tau + \alpha d\sin\omega\tau$ $d = \varepsilon^{-\alpha\tau}, r^2 = \alpha^2+\omega^2$
$\frac{p+\alpha}{(p+\alpha)^2+\omega^2}$	$\frac{\tau(\alpha\tau+2)}{g}\frac{[1,1+h/k,h/k]}{\left[1,-2f/g,1-\frac{8\alpha\tau}{g}\right]}$ $h = \alpha\tau - 2, k = \alpha\tau + 2$ f, g as above	$\frac{C}{2r^2}\frac{[1,1+D/C,D/C]}{[1,-2d\cos\omega\tau,d^2]}$ $C = \alpha - \alpha d\cos\omega\tau + \omega d\sin\omega\tau$ $D = \alpha d^2 - \alpha d\cos\omega\tau - \omega d\sin\omega\tau$

* Serial operator for inputs having zero initial value, see Section 3.2.

In the case of the simpler transfer functions the serial operators may be compared, as shown in Table 3. From this it can be seen that for $Y(p) = 1/p$, $1/p^2$, the same serial operators result in both methods. In the case of $Y(p) = 1/p^3$ the different operators $\frac{\tau^3}{8}\frac{[1,3,3,1]}{[1,-1]^3}$ and $\frac{\tau^3}{12}\frac{[1,5,5,1]}{[1,-1]^3}$ may be compared by integrating three times the ramp function [0, 1, 2, 3 . . .]. The results are [0, 0·125, 1·0, 4·13, 12·0, 28·13, 57, 104·1, 176·0 . . .] and [0, 0·0834, 0·834, 3·75, 11·33, 27·1, 55·5, 102·1, 173·4 . . .]. The correct result is [0, 0·0416, 0·667, 3·37, 10·67, 26·05, 54·0, 100·0, 170·6 . . .]. The second is the more effective operator.

expansion carried out, the result

$$\frac{1}{2}\left(\alpha\tau - \frac{\alpha^2\tau^2}{2} + \frac{\alpha^3\tau^3}{4}\right)\frac{[1,1]}{\left[1, -\left(1 - \alpha\tau + \frac{\alpha^2\tau^2}{2} - \frac{\alpha^3\tau^3}{4}\right)\right]}$$

is obtained. Since the two operators differ only in the cubi term in $\alpha\tau$ they will give the same result if a reasonable spacin of the ordinates is chosen in the first place, i.e. τ reasonably small It is difficult to state, in general, which of the two methods wil give the greater accuracy, since this appears to depend on th

type of inputs being considered. The method proposed here, however, is exact for step-function inputs. The serial operator for any transfer function is then given by the j.t.f. without modification by the factor $\frac{1}{2}(1 + z)$.

(4) CONCLUSION

The use of jump functions in approximate analysis has been demonstrated and the connection with serial numbers and serial operators has been discussed. The jump-transfer function has been defined as the ratio of the Laplace transforms of two jump functions. These are the input to a filter and the jump-function approximation of the actual output. The jump-transfer function may be replaced, for the purposes of approximate analysis, by an equivalent serial operator, and rules have been given for the formation of this operator for any transfer function. The method is exact for step-function inputs.

A "closed form" serial number representation of any input (except an impulse input) has been deduced from the Laplace transform of the jump-function representation of that input. This is, in effect, a summation process which may be done using the Laplace transform. The procedure has been demonstrated numerically and compared with existing methods of calculating approximate responses.

(5) ACKNOWLEDGMENT

The author is at present Edward A. Deeds, Fellow of University College, Dundee, in the University of St. Andrews, and he wishes to acknowledge his gratitude to Professor E. G. Cullwick for his generous encouragement.

(6) REFERENCES

(1) TUSTIN, A.: "A Method of Analysing the Behaviour of Linear Systems in terms of Time Series," *Journal I.E.E.*, 1947, **94**, Part IIA, p. 130.

(2) BROWN, B. M.: "Application of Finite Difference Operators to Linear Systems," *Automatic and Manual Control* (Butterworth, 1952), p. 409.

(3) BARKER, R. H.: "The Pulse Transfer Function and its Application to Sampling Servo-Systems," *Proceedings I.E.E.*, Monograph No. 43 M, July, 1952 (**99**, Part IV, p. 302).

(4) GARDNER, M. F., and BARNES, J. L.: "Transients in Linear Systems" (Wiley, 1942).

(5) LAWDEN, D. F.: "The Function $\sum_{n=1}^{\infty} n^r z^n$ and Associated Polynomials," *Proceedings of the Cambridge Philosophical Society*, 1951, **47**, p. 309.

621.391:621.3.018.7

Monograph No. 111 R
Nov. 1954

THE EFFECT OF SEVERE AMPLITUDE LIMITATION ON CERTAIN TYPES OF RANDOM SIGNAL: A CLUE TO THE INTELLIGIBILITY OF "INFINITELY" CLIPPED SPEECH

By J. M. C. DUKES, M.A., Associate Member.

(*The paper was first received 26th November*, 1953, *and in revised form 26th June*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in November*, 1954.)

SUMMARY

The paper is concerned with the effect of "infinite" peak clipping[1–3] on certain types of signal definable in statistical terms. The analysis is further extended to include subsequent differentiation of the clipped waveform with respect to time.

It is shown that with certain signals of common occurrence there is a marked similarity in spectral content before and after clipping. Furthermore, the phase relationships present a high degree of coherence.

Speech waveforms have certain attributes in common with these idealized signals, and it is therefore concluded that the high intelligibility of "infinitely" clipped speech is a phenomenon to be expected.

(1) INTRODUCTION

It has been found, as a result of experiment, that a large degree of non-linearity can be tolerated in an amplifier before there is any marked degradation in the intelligibility of normal speech. In the limit the amplifier may have a characteristic of the form shown in Fig. 1, where the amplitude difference OA

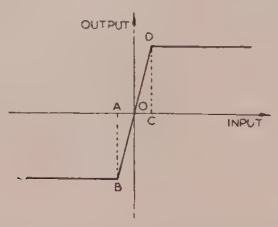


Fig. 1.—Amplifier characteristic.

is about 60 dB below the average level of the speech input signal. This is sometimes referred to as "infinite" peak clipping.[1] Assuming an input signal of the general form shown in Fig. 2(*a*), the output waveform of the amplifier will consist, for practical purposes, of a train of positive- and negative-going rectangular pulses of varying width, as in Fig. 2(*b*). For convenience we assume the clipping levels to be +1 and −1 units of amplitude.

If this clipped signal is now differentiated, we obtain the pulse train shown in Fig. 2(*c*). In practice, of course, the pulses will be of finite width and finite amplitude.

If either train of pulses is fed into a loudspeaker it will be found that, although the quality of the reproduced sound is very poor, the intelligibility is high.[1,2,3] More specifically, articulation tests based on sets of one-syllable words selected by a carefully controlled random process show articulation scores of the general order of 90% in each case. An initial period of accommodation is required whilst the listener gets used to the noisy character of the sound, and, moreover, fatigue is liable to set in fairly soon if the test is continued too long. By passing the

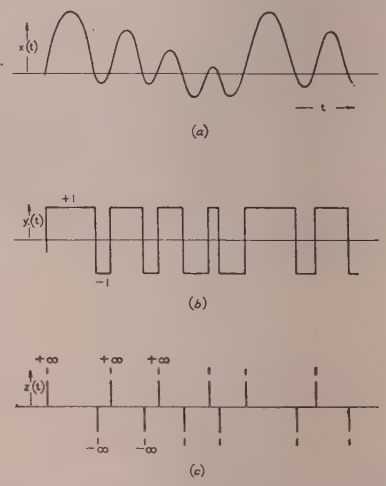


Fig. 2.—Effect of "infinite" peak clipping on a typical waveform.
(*a*) Original waveform.
(*b*) Clipped waveform.
(*c*) Clipped waveform with subsequent differentiation.

pulse trains through a low-pass filter, the tonal quality is improved, but the effect on the intelligibility is not important.

When comparing the distorted waveform with the original it is evident that the only characteristic that is retained is the relative position of the zeros of the signal. It would therefore seem logical to conclude that the greater part of the information-bearing content of speech must be contained in the relative positions of the zeros. This conclusion, however, seems to be at variance with the almost universally accepted concept of the ear as a short-time Fourier analyser.

The object of this paper, therefore, is to examine to what extent the spectral content of the waveforms of Figs. 2(*b*) and 2(*c*) is similar *on the average* to that of the original signal of Fig. 1(*a*). More important still, however, is the degree of coherence between the two spectra under consideration, i.e. the extent to which corresponding regions of the spectrum are phase-related in a fixed rather than a random manner. (The term "coherence" is defined precisely in Section 3.3, but its use in this context is by no means new.[4])

If it can be shown, as in fact it will be, that the spectral energy distribution is to a large extent invariant under clipping, it may be concluded that the high intelligibility of clipped speech is a phenomenon to be expected. However, it should be stressed that the method is only valid in so far as it relates to averages over long periods of time. Further and more detailed work is still required to show what are the important invariants in the case of individual sounds.

In order to simplify the analysis, various assumptions will be made regarding the character of the original waveform. It

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Dukes is with Standard Telecommunication Laboratories, Ltd.

will be shown that these simplifications do not in fact invalidate the essential conclusions in any way.

(2) REPRESENTATION OF THE ORIGINAL SIGNAL

In order to simplify the analysis that follows the continuous waveform of Fig. 2(*a*) will be time quantized in the manner illustrated in Fig. 3(*a*). The more important results of the paper

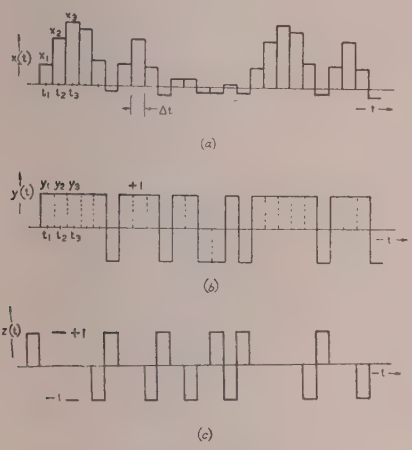


Fig. 3.—Time-quantized waveforms.
(*a*) Original pulse train [time-quantized version of Fig. 2(*a*)].
(*b*) Clipped pulse train.
(*c*) Differentiated clipped pulse train, with pulse reshaping.

are independent of the length of the quantizing interval, which, if so desired, may be considered as tending to zero.

As regards the signal amplitudes, it does not matter whether we assume the range of values to be continuous or to be quantized in discrete levels. From the notational point of view it is simpler to assume continuous functions, but this does not in any way constitute a restriction.

The waveform of Fig. 3(*a*) will be referred to throughout the paper as *the original pulse train*, and that of Fig. 3(*b*) as the *clipped pulse train.*

The original pulse train is assumed to be generated by some form of stationary random processes;[5,6] i.e. the statistics of the signal remain invariant under any shift in the origin of time. No consideration is given to purely periodic waveforms, since these may be handled by simpler and more conventional techniques.[7,8]

For the purpose of this analysis, signals will be divided into two classes as follows:

(*a*) Signals produced by a generating process such that the amplitude of one pulse is in no way modified by the nature of the pulse sequence immediately preceding that pulse. This will be referred to as a *totally random signal.*

(*b*) Signals such that the probability of a given amplitude is a function also of the sequence of amplitude levels preceding the pulse under consideration. This will be referred to as a *partially constrained signal.*

For the most part the paper deals with signals in category (*a*). These may be defined simply by means of the one-dimensional probability distribution $p(x)$, where $p(x)\delta x$ is the probability that the amplitude x of a pulse lies between x and $x + \delta x$.

No restrictions are placed on the form of $p(x)$ except that

$$\int_{-\infty}^{+\infty} p(x)dx = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

thus normalizing the distribution $p(x)$.

$$\int_{-\infty}^{+\infty} xp(x)dx = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

i.e. the d.c. level of the original pulse train is zero.

$$\int_{-\infty}^{0} p(x)dx = \int_{0}^{+\infty} p(x)dx \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The third restriction, which in effect states that positive and negative amplitudes are equally likely, ensures that there is no d.c. component in the clipped pulse train. The effect of relaxing this restriction is considered in Section 6.

No other restrictions are placed on the form of $p(x)$ except, of course, that it must be wholly real. For amplitude quantization, $p(x)$ will consist of a set of discrete values instead of a continuous function, but in the present instance this provokes no special difficulties.

With the partially constrained signal it is necessary to make use of the n-dimensional probability distribution $p(x_r, x_s)$, which denotes the joint probability that x lies between x_r and $x_r + \delta x$ at time t_r, and between x_s and $x_s + \delta x$ at time t_s.

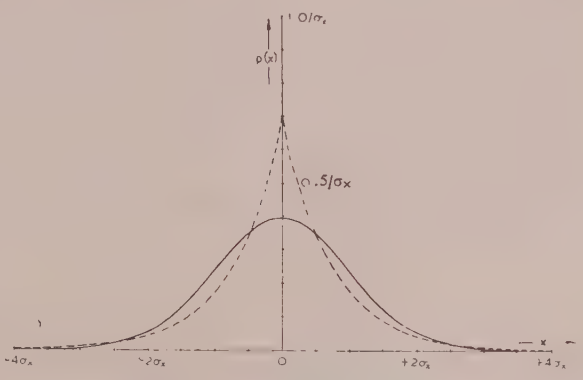


Fig. 4.—Probability distributions.

——— $p_1(x) = \frac{1}{\sigma_x\sqrt{(2\pi)}}\varepsilon^{-x^2/2\sigma_x^2}$

— — — $p_2(x) = \frac{1}{\sigma_x\sqrt{2}}\varepsilon^{-\sqrt{2}|x|/\sigma_x}$

Two specific probability distributions for $p(x)$ will be considered, both of which satisfy the above restrictions, namely

$$p_1(x) = \frac{1}{\sigma_x\sqrt{(2\pi)}}\varepsilon^{-x^2/2\sigma_x^2} \quad . \quad . \quad . \quad . \quad . \quad (4)$$

and

$$p_2(x) = \frac{1}{\sigma_x\sqrt{2}}\varepsilon^{-\sqrt{(2)}|x|/\sigma_x} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where σ_x is the standard deviation, defined as

$$\sigma_x = \left[\int_{-\infty}^{+\infty} x^2 p(x)dx\right]^{\frac{1}{2}} \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The first is the Gaussian distribution and is characteristic of thermal noise. Moreover, it has been found experimentally that it also represents, to a close approximation, the instantaneous amplitude distribution of unvoiced speech sounds[9] (e.g. fricatives). The second, sometimes referred to as the exponential distribution, is a close approximation to the instantaneous

amplitude distribution of voiced speech sounds[9] (e.g. vowels, semi-vowels, etc.).

True speech signals should in fact be classified as partially constrained signals. However when $t_s \to t_r$, $p(x_r, x_s)$ tends to a single probability distribution $p(x)$. Hence it will be evident that under certain circumstances it will be legitimate to use eqns. (4) and (5) to derive results relating to speech intelligibility, and this in fact has been done.

(3) SOME PROPERTIES OF THE CORRELATION FUNCTION

(3.1) Definition of the Correlation Function[5,6]

Let the original signal (or pulse train) be represented by the function $x(t)$, and the distorted signal (or pulse train) by $y(t)$. The cross-correlation function may be expressed as

$$R_{xy}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} x(t)y(t+\tau)dt \quad . \quad . \quad . \quad (7)$$

where τ is the time displacement of the signal $y(t)$ relative to the signal $x(t)$.

It is customary to normalize the correlation function in the following manner:

$$\rho_{xy}(\tau) = R_{xy}(\tau)/\sigma_x\sigma_y \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

with the result that the function now ranges between limits of $+1$ and -1.

If the function $y(t)$ is identical to the function $x(t)$, it is known as the auto-correlation function.

When the signals $x(t)$ and $y(t)$ are each generated by a stationary random process[5,6] we are obliged to redefine eqn. (8) in statistical terms as follows:

$$\rho_{xy}(\tau) = \frac{1}{\sigma_x\sigma_y} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyp(x, y : \tau)dxdy \quad . \quad . \quad . \quad (9)$$

where $p(x,y : \tau)$ denotes the joint probability of x lying between x and $x + \delta x$ and y lying between y and $y + \delta y$ when the displacement of $y(t)$ relative to $x(t)$ is τ.

It will be observed that whereas eqn. (8) represents a time average of the product of the two signals expressed as functions of time, eqn. (9) is an average over the ensemble of possible product values at a single instant of time. The use of the second form instead of the first is legitimate only when stationary random conditions prevail.

(3.2) The Correlation Function and the Spectral-Energy-Density Function

It has been shown elsewhere[5,6] that the correlation function and the spectral-energy-density function are a Fourier transform pair,

i.e. $$\rho(\tau) = \int_{-\infty}^{+\infty} g(f)\varepsilon^{+2j\pi f\tau}df \quad . \quad . \quad . \quad . \quad (10a)$$

and $$g(f) = \int_{-\infty}^{+\infty} \rho(\tau)\varepsilon^{-2j\pi f\tau}d\tau \quad . \quad . \quad . \quad . \quad (10b)$$

where $g(f)$ is the spectral-energy-density function, and both $\rho(\tau)$ and $g(f)$ are normalized functions.

In the case where $\rho(\tau)$ is an auto-correlation function the above result may be verified without much difficulty. However, when $\rho(\tau)$ is a cross-correlation function, and $g(f)$ is therefore a cross-spectral-energy-density function, the derivation is a little less obvious. For this reason, and because of the dominant role that relationship plays in the subsequent analysis, it is considered advisable to outline briefly the manner in which it may be derived.

In the first place it should be emphasized that a random signal, like other signals, may be defined by a complex spectrum $a_x(f)$. (Once again this is assumed to be a normalized function.) However, whereas $|a_x(f)|$ is determinate, and hence also $g_x(f)$, the phase function $\phi_x(f)$ can only be expressed in statistical terms. It will therefore be evident that to derive the cross-spectral-energy-density function we must of necessity average the product $a_x^*(f)a_y(f)$ over all possible values of $(\phi_x - \phi_y)$. Moreover, whereas $g_{xx}(f)$ must be wholly real, $g_{xy}(f)$ may in the general case contain both real and imaginary parts, each of which is an independent function of frequency.

Since $|a_x(f)|$ and $|a_y(f)|$ are determinate, so also must be the product $|a_x(f)||a_y(f)|$. Hence all we require is the average value of $\varepsilon^{+j[\phi_y(f)-\phi_x(f)]}$ as a function of frequency f. Without any formal analysis it is easy to show by means of a diagram such as Fig. 5 that the average value takes the form

$$K(f)\varepsilon^{-j\Phi_{xy}(f)}$$

where $|K(f)| \leqslant 1$ and both $K(f)$ and Φ_{xy} are determined by the probability distributions appropriate to $\phi_x(f)$ and $\phi_y(f)$ and the properties of the filter.

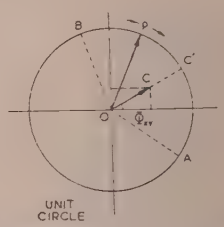


Fig. 5.—Phase fluctuation on the Argand diagram.
P fluctuates between A and B.
C is the mean position of P.
Φ_{xy} is the mean value of $(\phi_x - \phi_y)$.

Hence the solution for $g_{xy}(f)$ is

$$g_{xy}(f) = |a_x(f)||a_y(f)|K(f)\varepsilon^{+j\Phi_{xy}(f)} \quad . \quad . \quad (11)$$

If the function $y(t)$ is displaced by τ relative to the function $x(t)$ we may rewrite this equation as

$$g_{xy}(f)_\tau = |a_x(f)||a_y(f)|K(f)\varepsilon^{+j[\Phi_{xy}(f)+2\pi f\tau} \quad . \quad (12)$$

By integrating eqn. (12) over all positive and negative values of f we obtain the average cross-product as a function of τ. But, by definition, this is in fact the cross-correlation function.

Hence $$\rho_{xy}(\tau) = \int_{-\infty}^{+\infty} g_{xy}(f)_\tau df \quad . \quad . \quad . \quad . \quad . \quad (13a)$$

$$= \int_{-\infty}^{+\infty} g_{xy}(f)\varepsilon^{+2j\pi f\tau}df \quad . \quad . \quad . \quad (13b)$$

which demonstrates the validity of eqn. (10a).

(3.3) Coherence

Although $g_{xy}(f)$ in eqn. (12) may contain both real and imaginary parts, the imaginary part must be odd, since $a_x(f)$ and $a_y(f)$ are the transforms of real functions of time, and

therefore integrate to zero when the integration is carried out over the full range of both positive and negative frequencies.

Hence
$$\rho_{xy}(\tau) = \int_{-\infty}^{+\infty} \mathscr{R}[g_{xy}(f)_\tau]df \quad . \quad . \quad . \quad . \quad (14)$$

Thus the correlation coefficient $\rho_{xy}(\tau_1)$ is a measure of the average in-phase energy of the two signals for a given time displacement τ_1. In other words it is a measure of the coherence of the signals.[4]

Now in general there will be some special value of τ, say τ', which makes $|\rho_{xy}(\tau)|$ an absolute maximum. This maximum value of the correlation coefficient will be called the "first coherence coefficient" μ;

i.e.
$$\mu = |\rho_{xy}(\tau')| \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15a)$$
$$= \left| \int_{-\infty}^{+\infty} g_{xy}(f)_{\tau'} df \right| \quad . \quad . \quad . \quad . \quad (15b)$$

In the case where $x(t)$ is the input to, and $y(t)$ the output from, a non-linear filter it is evident that μ may under certain circumstances provide a useful first-order measure of the expected average intelligibility at the output. On the other hand it is possible that a better measure will be given by

$$\nu = \int_{-\infty}^{+\infty} |g_{xy}(f)| df \quad . \quad . \quad . \quad . \quad . \quad (16)$$

which will be referred to as the "second coherence coefficient" [note that $|g_{xy}(f)|$ is independent of τ].

The difference between the two coefficients may be clarified by a specific example. For simplicity let us suppose that $y(t)$ is derived from $x(t)$ by means of a passive all-pass network having a phase characteristic that is a non-linear function of frequency. Evidently in this case μ must of necessity be numerically less than unity, because there is no single value of τ which will ensure that the spectrum of $y(t+\tau)$ is in phase with the spectrum of $x(t)$ throughout the entire frequency range. On statistical grounds, however, these two signals might be considered as being coherent, and this viewpoint would be in accordance with the observation that the ear is relatively insensitive to phase. Furthermore, it is evident that there always exists, in theory at any rate, an inverse network capable of recovering the original signal in its entirety.

The above considerations suggest that when a non-linear filter incorporates frequency-sensitive elements the second coherence coefficient is likely to prove a more satisfactory indication of intelligibility.

(4) THE EFFECT OF INFINITE CLIPPING ON A TOTALLY RANDOM SIGNAL

(4.1) The Cross-Correlation Function

The original waveform before clipping is illustrated in Fig. 3(*a*) and has the mathematical properties outlined in Section 2. The clipped waveform, shown in Fig. 3(*b*), will be symbolized by the function $y(t)$. To a set of values x_1, x_2, x_3, etc., at t_1, t_2, t_3, etc., there corresponds a set of values y_1, y_2, y_3, etc. By definition only two values of y are possible, namely $+1$ or -1.

The mechanism of clipping may be symbolized mathematically by the relationship

$$y_r = x_r^{1/n} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

where n is a very large odd integer. In the limit, n may be assumed to tend to infinity.

To evaluate the cross-correlation function of the two signals use will be made of eqn. (9). However, it is necessary first to obtain expressions for the probability distribution $p(x, y; \tau)$, which may be done in the following manner.

Consider the situation relevant to one pair of channels out of the ensemble of pairs of channels. This is illustrated in Fig. 6,

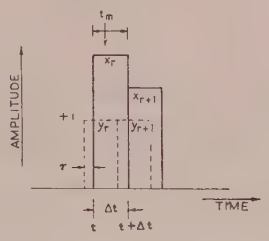


Fig. 6.—Calculation of the correlation function $\rho_{xy}(\tau)$.

where t_m is the instant of time at which the observation is made. Let this observation be made during the rth pulse. Evidently t_m may lie anywhere between t and $t + \Delta t$ with equal probability. Hence provided $|\tau| < \Delta t$ we may say that the probability of an overlap of x_r with y_r at t_m is $(\Delta t - |\tau|)/\Delta t$, and the probability of an overlap of x_r with y_{r+1} at t_m is $|\tau|/\Delta t$. However, given that x_r and y_r do overlap, the joint probability $p(x_r, y_r)dxdy$ is identical to the probability $p(x_r)dx$, whilst given that x_r and y_{r+1} overlap, the joint probability $p(x_r, y_{r+1})dxdy$ is identical to the probability $p(x_r)p(x_{r+1})dxdy$, i.e. the product of the independent probabilities.

Hence for the range $|\tau| < \Delta t$ we may rewrite eqn. (9) in the form

$$\rho_{xy}(\tau) = \left[\frac{\Delta t - |\tau|}{\Delta t} \int_{-\infty}^{+\infty} x_r^{1+1/n} p(x_r) dx + \frac{|\tau|}{\Delta t} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_r x_{r+1}^{1/n} p(x_r) p(x_{r+1}) dx dx \right] \Big/ \sigma_x \sigma_y \quad . \quad . \quad . \quad . \quad (18)$$

This expression may be simplified by noting that

(*a*) The second integral equates to zero.

(*b*) Since n is odd, $x_r^{1+1/n} = |x_r|$.

(*c*) As a result of the restriction involved in eqn. (3)

$$\int_{-\infty}^{+\infty} |x_r| p(x_r) dx = -2 \int_{-\infty}^{0} x_r p(x_r) dx = +2 \int_{0}^{+\infty} x_r p(x_r) dx \quad . \quad (19)$$

To simplify the notation, the quantity $\int_0^{+\infty} x_r p(x_r) dx$, which represents the average value of the positive-going pulses (or alternatively the average value of the negative-going pulses), will be given the symbol A_x.

By noting furthermore that

(*d*) The r.m.s. level of $y(t)$, i.e. σ_y, is unity.

(*e*) For $|\tau| > \Delta t$ all integrals equate to zero [for the same reasons as does the second integral of eqn. (18)], we finally obtain the result that

$$\rho_{xy}(\tau) = \frac{2A_x}{\sigma_x} \frac{(\Delta t - |\tau|)}{\Delta t} \quad \text{for } |\tau| < \Delta t$$
$$= 0 \quad \text{for } |\tau| > \Delta t \quad . \quad . \quad (20)$$

This is illustrated graphically in Fig. 7.

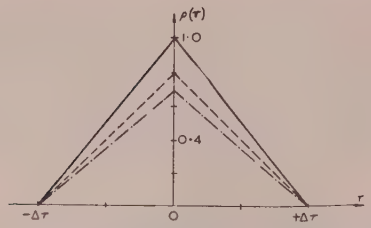


Fig. 7.—Normalized correlation functions for clipped pulse train.
——— $\rho_{xx}(\tau)$ and $\rho_{yy}(\tau)$: Auto-correlation functions before and after clipping.
— — — $\rho_{xy}(\tau)$: Cross-correlation function for distribution $p_1(x)$.
—·—·— $\rho_{zy}(\tau)$: Cross-correlation function for distribution $p_2(x)$.

The first coherence coefficients for the two specific probability distributions considered in Section 2 may now be evaluated as follows:

$$\mu_{xy} = 2A_x/\sigma_x = \sqrt{(2/\pi)} = 0{\cdot}798, \text{ for } p_1(x) \quad . \quad (21)$$

$$\mu_{xy} = 2A_x/\sigma_x = 1/\sqrt{2} = 0{\cdot}707, \quad \text{for } p_2(x) \quad . \quad (22)$$

It is interesting to note how high these values are numerically. It is also interesting that they are generally of the same order as the percentage articulation scores obtained during tests with clipped speech. The fact that articulation scores average some 10% higher is not surprising, however, in that the above analysis is pertinent only to a totally random signal. This point will be considered again in Sections 7 and 9.

(4.2) The Auto-Correlation Functions

Applying the same methods as before, the following interesting result is obtained:

$$\rho_{xy}(\tau) = \mu_{xy}\rho_{xx}(\tau) = \mu_{xy}\rho_{yy}(\tau) \quad . \quad . \quad . \quad (23)$$

In other words, the two auto-correlation functions and the cross-correlation function are identical in shape. The absolute amplitudes differ only by a constant factor of proportionality equal to the first coherence coefficient.

(4.3) The Spectral-Energy-Density Functions

Since the three correlation functions are identical except for a constant factor of proportionality, it is evident that the three spectral-energy-density functions will also be identical except for the same constant factor. Applying the transform given in eqn. (10*b*) to eqn. (20) the following result is obtained:

$$g_{xx}(f) = g_{yy}(f) = \frac{\sin^2\pi f\Delta t}{\pi^2 f^2 \Delta t} \quad . \quad . \quad . \quad (24a)$$

and

$$g_{xy}(f) = \mu g_{xx}(f) \quad . \quad . \quad . \quad . \quad . \quad (24b)$$

Since $g_{xy}(f)$ is wholly real for all values of f, it follows that both definitions of coherence are identical;

i.e. $$\nu = \mu$$

The spectral-density functions are plotted in Fig. 8 for values of μ appropriate to the Gaussian and exponential distributions of eqns. (4) and (5).

(5) DIFFERENTIATED CLIPPED RANDOM PULSE TRAIN

(5.1) Outline of the Method

If the infinitely clipped waveform of Fig. 2(*b*) is differentiated, a train of alternate positive and negative delta impulses will be obtained as in Fig. 2(*c*). These impulses are theoretically of zero duration and infinite amplitude, but in practice they will be smoothed by the finite bandwidth of the system. As an analytical

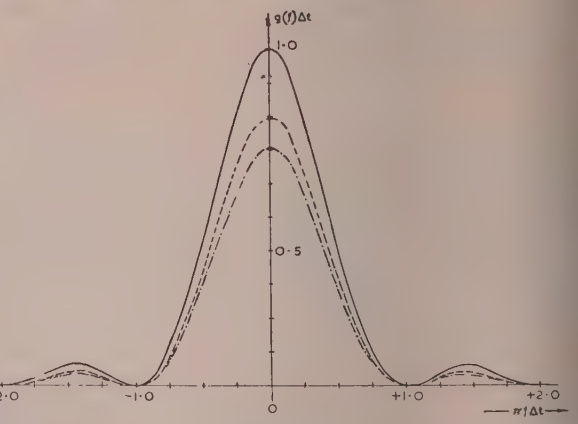


Fig. 8.—Normalized spectral-energy-density functions for clipped pulse train.
——— $g_{xx}(f), g_{yy}(f)$: Spectral-energy-density functions before and after clipping.
— — — $g_{xy}(f)$: Cross-spectral-energy-density function for $p_1(x)$.
—·—·— $g_{zy}(f)$: Cross-spectral-energy-density function for $p_2(x)$.

simplification it is assumed that they are reshaped to form unit-amplitude rectangular pulses of the same duration Δt as the original time-quantized pulse train [see Fig. 3(*c*)].

In comparing Fig. 3(*c*) with Fig. 3(*b*) it is evident that the possible signal transformations may be grouped into four main categories as follows:

(*a*) Two positive values of $x(t)$ in succession. This produces zero output after differentiation; hence $z(t) = 0$.
(*b*) Two successive negative values of $x(t)$. Again $z(t) = 0$.
(*c*) A positive value of x is succeeded by a negative value of x. In this case $z = -1$.
(*d*) A negative value of x is succeeded by a positive value of x. Hence $z = +1$.

It is not difficult to show that the absolute probability of each category is $\frac{1}{4}$.

The analysis then proceeds on lines somewhat similar to those of Section 4.1. This involves a calculation of the probabilities of the two possible products which may occur for each of the four signal sequences described above. These probabilities will be functions of the time displacement τ. Multiplying the product by its probability and integrating over the ensemble gives the average product for each category. The sum of the four average products, each weighted by $\frac{1}{4}$, expressed as a function of τ, is by definition the correlation function.

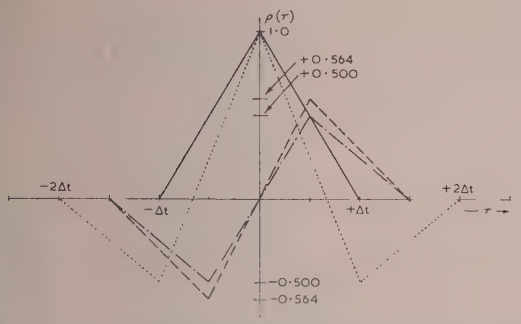
Full details of this analysis are given elsewhere.[7] However, it should be noted a considerable simplification may be effected by making use of the principles outlined in Section 7.2.

The final result is that

$$\left.\begin{aligned}
\rho_{xz}(\tau) &= -\sqrt{(2)}A_x(3\Delta t + 2\tau)/(2\sigma_x\Delta t), \\
&\qquad \text{for } -\tfrac{3}{2}\Delta t < \tau < -\tfrac{1}{2}\Delta t \\
&= +2\sqrt{(2)}A_x\tau/(\sigma_x\Delta t), \\
&\qquad \text{for } -\tfrac{1}{2}\Delta t < \tau < +\tfrac{1}{2}\Delta t \\
&= +\sqrt{(2)}A_x(3\Delta t - 2\tau)/(2\sigma_x\Delta t), \\
&\qquad \text{for } +\tfrac{1}{2}\Delta t < \tau < +\tfrac{3}{2}\Delta t \\
&= 0 \text{ for all other values of } \tau
\end{aligned}\right\} \quad . \quad (25)$$

This is illustrated in Fig. 9.
By inspection it is evident that

$$\mu_{xz} = \sqrt{(2)}A_x/\sigma_x \quad . \quad . \quad . \quad . \quad . \quad (26$$



ig. 9.—Normalized correlation functions for differentiated clipped pulse train, with pulse reshaping.

——— $\rho_{xz}(\tau)$: Auto-correlation function of $x(t)$.
..... $\rho_{zz}(\tau)$: Auto-correlation function of $z(t)$.
— — — $\rho_{xz}(\tau)$: Cross-correlation function for distribution $p_1(x)$.
—·—·— $\rho_{xz}(\tau)$: Cross-correlation function for distribution $p_2(x)$.

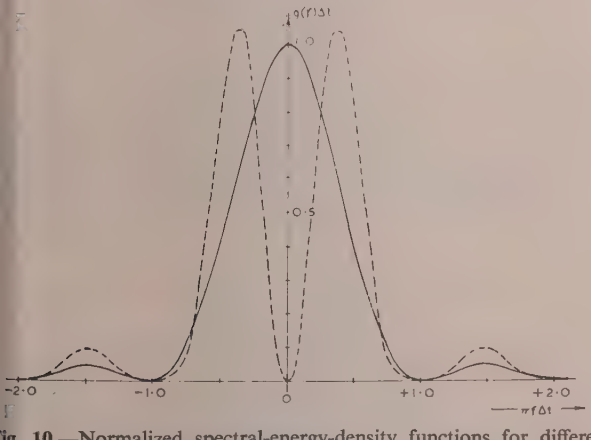
Comparing this result with that obtained for the clipped pulse rain it can be seen that the effect of differentiation has been to educe the first coherence coefficient by a factor of $\sqrt{2}$. The ignificance of this is discussed in Section 5.3.

(5.2) The Auto-Correlation Function of $z(t)$

Computation of the auto-correlation function of the differentiated clipped signal is much simplified by the method of quantization outlined in Section 7.2. The result is illustrated in Fig. 9(*b*).

(5.3) The Cross-Spectral-Energy-Density Function

The Fourier transform of the cross-correlation function may e evaluated by standard methods without difficulty. The



ig. 10.—Normalized spectral-energy-density functions for differentiated clipped pulse train, with pulse reshaping.

——— $g_{xx}(f)$: Spectral function for original pulse train.
— — — $g_{zz}(f)$: Spectral function for differentiated clipped pulse train with pulse reshaping.

llowing result for the cross-spectral-energy-density function is btained:

$$g_{xz}(f)\tau' = -2j\frac{\sqrt{(2)}A_x \sin^3 \pi f\Delta t}{\sigma_x \pi^2 f^2 \Delta t}\varepsilon^{+j\pi f\Delta t} \quad . \quad . \quad . \quad (27a)$$

$$= \frac{2\sqrt{(2)}A_x}{\sigma_x}\left(\frac{\sin^4 \pi f\Delta t}{\pi^2 f^2 \Delta t} - j\frac{\sin^3 \pi f\Delta t \cos \pi f\Delta t}{\pi^2 f^2 \Delta t}\right)$$

$$. \quad . \quad . \quad . \quad (27b)$$

here $\tau' = \Delta$

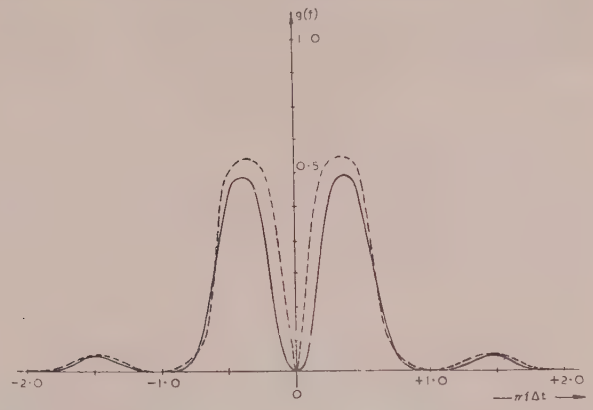


Fig. 11.—Cross-spectral-energy-density functions for differentiated clipped pulse train, with pulse reshaping.

— — — Modulus of $g_{xy}(f)$.
——— Real part of $g_{xy}(f)$ for a relay $\tau = \pm\frac{1}{2}\Delta t$.
Curves are for distribution $p_1(x)$. For distribution $p_2(x)$, multiply ordinate values by 0·89.

The real part of eqn. (27*b*) has been plotted in Fig. 11 for the first probability distribution [see eqn. (4)].

The second coherence coefficient may be written

$$\nu_{xz} = \int_{-\infty}^{+\infty} |g_{xz}(f)|\,df \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

$$= \frac{2\sqrt{(2)}A_x}{\sigma_x}\int_{-\infty}^{+\infty}\left|\frac{\sin^3 \pi f\Delta t}{\pi^2 f^2 \Delta t}\right| df \quad . \quad . \quad . \quad (28)$$

The solution to the above integral, suggested by L. Lewin, leads to the result that

$$\nu_{xz} = 4\sqrt{(2)}A_x/\pi\sigma_x = 1{\cdot}273\ \mu_{xz}$$
$$= 0{\cdot}900\ \mu_{xy} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

Hence the effect of differentiating the clipped signal may be summarized as follows:

(*a*) The first coherence coefficient is reduced by approximately 29%.
(*b*) The second coherence coefficient is reduced by only 10%.

The latter result is of interest in that it has been found that subsequent differentiation does not materially affect the intelligibility of clipped speech.

(5.4) The Spectral-Energy-Density Function of $z(t)$

There are three possible ways of obtaining the spectral-energy-density function $g_{zz}(f)$ of the differentiated clipped signal:

(*a*) By computing the Fourier transform of $\rho_{zz}(\tau)$.
(*b*) By substituting $g_{xz}(f)_{\tau'}$ and $|a_x(f)| = [g_{xx}(f)]^{\frac{1}{2}}$ in eqn. (12) and identifying the correct expressions for $K(f)$ and $\Phi_{xz}(f)$.
(*c*) By operating on the expression for $g_{yy}(f)$ given in eqn. (24*a*).

The first is rather tedious. The second is instructive and leads to the result $K(f) = 2A_x/\sigma_x$ and $\Phi_{xz}(f) = -\pi/2$. The third method, however, is the simplest and will therefore be described.

Differentiation is equivalent to passing the signal through a filter whose response rises with frequency at the rate of 6dB per octave. Hence the first step is to multiply $g_{yy}(f)$ by $(2\pi f\Delta t)^2$. To reshape the signal into a rectangular pulse train we imagine it to be passed through a filter having a $\sin x/x$ type response. This in turn involves multiplication by $(\sin \pi f\Delta t)^2/\pi^2 f^2 \Delta t$.

Finally, in order to normalize the result we multiply by a suitable constant so as to satisfy the requirement

$$\int_{-\infty}^{+\infty} g_{zz}(f)df = 1 \quad . \quad . \quad . \quad . \quad . \quad (30)$$

This leads to the simple result

$$g_{zz}(f) = 2(\sin \pi f \Delta t)^4/\pi^2 f^2 \Delta t \quad . \quad . \quad . \quad (31)$$

By comparing eqns. (27) and (31) it will be seen that

$$g_{zz}(f) = \mathcal{R}[g_{xz}(f)]/\mu_{xz} \quad . \quad . \quad . \quad . \quad (32)$$

(6) THE EFFECT OF ASYMMETRY IN THE ORIGINAL WAVEFORM

The third restriction on the form of $p(x)$, given in eqn. (3) of Section 2 will now be relaxed. As already pointed out, the clipped signal will then possess a d.c. component.

Let $$\int_{-\infty}^{0} p(x)dx = k\int_{0}^{+\infty} p(x)dx \quad . \quad . \quad . \quad . \quad (33)$$

where $$0 < k < +\infty$$

Then the probability of a positive pulse is $1/(1 + k)$ and the probability of a negative pulse is $k/(1 + k)$.

If $\bar{y}$ is the average value of y, or in other words the d.c. component, then

$$\bar{y} = (1 - k)/(1 + k) \quad . \quad . \quad . \quad . \quad . \quad (34)$$

and

$$\sigma_y = (\overline{y^2} - \bar{y}^2)^{\frac{1}{2}}$$

$$= 2\sqrt{(k)}/(1 + k) \quad . \quad . \quad . \quad . \quad (35)$$

The presence of a d.c. component does in principle necessitate the substitution of $(y - \bar{y})$ for y in eqns. (7) and (9). However, in the present instance $\bar{x}$ is zero. This being so, it may easily be shown that the result is independent of whether or not the substitution is made. Hence it follows that the only modification is due to a change in value of σ_y. The first coherence coefficient is thus increased to

$$\mu_{xy} = A_x(1 + k)/\sigma_x\sqrt{k} \quad . \quad . \quad . \quad . \quad (36)$$

This result, which may seem a little surprising at first sight, can conveniently be demonstrated by reference to Fig. 12. As

Fig. 12.—Probability distributions with increasing asymmetry.

the degree of asymmetry is progressively increased, the tendency is to concentrate the probability curve into a narrow thin hump of high probability on one side of zero, balanced by a much wider range of possible values, all of relatively low probability, on the other side. The average product is accordingly increased because negative values of y are much more frequent than positive values, and the set of possible values of x producing a negative value of y are much more restricted in range.

In the case of the differentiated clipped signal the following considerations apply. First there can be no d.c. compone after differentiation. Secondly, the probabilities of the sign transformation categories (*c*) and (*d*) described in Section are each reduced from $\frac{1}{4}$ to $k/(1 + k)^2$. Thirdly, σ_z is reduc from $1/\sqrt{2}$ to $2\sqrt{(k)}/(1 + k)$.

Hence the first coherence coefficient is reduced to

$$\mu_{xz} = 2\sqrt{(2)}A_x\sqrt{(k)}/\sigma_x(1 + k). \quad . \quad . \quad (3$$

The reason why this result is the contrary of the previo result is to be found in the operation of the assumed puls reshaping circuit which follows the differentiating circuit. trace of the original asymmetry is lost, and therefore the fir waveform becomes less and less like the original as the asy metry is increased.

In conclusion, it should be emphasized that relaxing restricti (*c*) in no way alters the shapes of either the correlation functio or the spectral-density functions. As has already been point out, these are independent of the form of $p(x)$.

(7) PARTIALLY CONSTRAINED SIGNALS

(7.1) Coherence

The general form of the correlation function is given eqn. (9). Letting $\tau \to 0$, it is evident that

$$p(x, y; \tau) \to p(x, y) \quad . \quad . \quad . \quad . \quad . \quad (3$$

But in the case of the clipped signal, y is a function of x onl and hence

$$p(x, y) \equiv p(x) \quad . \quad . \quad . \quad . \quad . \quad (3$$

Hence $\rho_{xy}(0)$ is a function only of the instantaneous amplitu distribution $p(x)$ and is independent of the transition probabilitie

But the first coherence coefficient is, by definition,

$$\mu = |\rho_{xy}(\tau')|$$

$$= |\rho_{xy}(0)| \text{ for the clipped signal.}$$

Hence with infinite clipping the first coherence coefficient determined entirely by the form of $p(x)$ and is independent the statistical constraints between successive signal values. the other hand, when y is a function of time also, as with t differentiated clipped signal, the coherence coefficient becomes function of the transition probabilities between successive sign values.

(7.2) Computation of the Correlation and Spectral-Density Functions

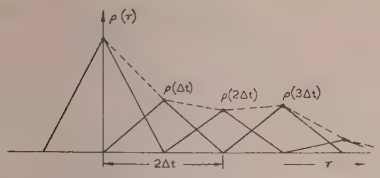
The following discussion is applicable whatever the form the non-linear filter producing y from x, provided that y is function of x only, and is therefore not limited to the simp case of infinite clipping. Certain of the observations can, hov ever, be applied to the case where there is time dependence, b a certain caution should be exercised.

In deriving eqn. (18) it was possible to assume that

$$p(x_r, y_{r+1}) = p(x_r)p(y_{r+1})$$

In the more general case of partially constrained signals it is n possible to make this simplification. Hence the second integr in eqn. (18) does not necessarily equate to zero. Furthermor for $|\tau| > \Delta t$, $\rho_{xy}(\tau)$ may have values other than zero.

It can be shown without much difficulty that, if the origin signal is time-quantized into rectangular pulses, as in Fig. 3(a the correlation functions can also be quantized into a series overlapping triangular pulses as illustrated in Fig. 13. The su of these pulses, denoted by the dotted line, defines the correlatio



g. 13.—Time-quantization of correlation function into elementary triangles.

nction. Hence it is only necessary to calculate $\rho(\tau)$ at the ecific values $\tau = 0, \pm\Delta t, \pm 2\Delta t$, etc.

This proposition can, without formal proof, be seen to follow a general manner from the nature of eqn. (18). The first tegral defines the triangular pulse centred at $\tau = 0$, and the cond integral (assuming it is non-zero) describes the triangular ılse centred at $\tau = \Delta t$. When τ exceeds Δt the first integral excluded from the equation, but a second one takes its place scribing the pulse centred on $\tau = 2\Delta t$. This process is repeated definitely for increasing values of τ.

The values of $\rho(\tau)$ at the points $m\Delta t$, where m is a positive or gative integer, or zero, are given by

$$\rho_{xy}(m\Delta t) = \frac{1}{\sigma_x\sigma_y}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy_m p(x, y_m)dxdy \quad . \quad (41)$$

here $y_m \equiv y(t + m\Delta t)$, if $x \equiv x(t)$

Now $$p(y_m, x) = p(y_m|x)p(x) \quad . \quad . \quad . \quad . \quad . \quad (42)$$

here $p(y_m|x)$ is the dependent probability distribution defining e probability of y at $(t + m\Delta t)$ given the value of x at t. This uation can be rewitten as

$$\rho_{xy}(m\Delta t) = \int_{-\infty}^{+\infty} xp(x)\left[\int_{-\infty}^{+\infty} y_m p(y_m|x)dy_m\right]dx \quad . \quad (43)$$

In general this represents a more useful formulation than jn. (41) in that the variables have to some extent been separated. urthermore, the expressions $p(x)$ and $p(y_m|x)$ can be deduced xperimentally in a somewhat more direct manner.

The same methods can be extended to the calculation of the ectral-density functions by virtue of the well-known Fourier perposition theorem.

hus $$g_{xy}(f) = \frac{\sin^2 \pi f\Delta t}{\sigma_x\sigma_y\pi^2 f^2\Delta t}\sum_m \rho_{xy}(m\Delta t)\varepsilon^{-j\pi fm\Delta t} \quad . \quad (44)$$

Finally it should be stressed that the above results are appli- ble to the study of both auto- and cross-correlation functions id both auto- and cross-spectral density functions. Thus to fine the auto-correlation function $\rho_{xx}(m\Delta t)$ it is only necessary eqn. (43) to substitute x_m for y_m.

(7.3) Relationship between the Three Correlation Functions

In the case of totally random signals it was shown that the ree correlation functions $\rho_{xx}(\tau)$, $\rho_{yy}(\tau)$, $\rho_{xy}(\tau)$ were all identical cept for a constant factor of proportionality. With par- lly constrained signals this is no longer necessarily true. hus for a sinusoidal input $x = \cos \omega t$, $\rho_{xx}(\tau)$ is sinusoidal in rm, and so is $\rho_{xy}(\tau)$, but $\rho_{yy}(\tau)$ is triangular.[7,8]

It would be interesting and perhaps useful to find the con- tions under which the following identities are either jointly or parately satisfied:

$$\rho_{xy}(\tau) = c\rho_{xx}(\tau) \quad . \quad . \quad . \quad . \quad . \quad . \quad (45)$$

$$\rho_{yy}(\tau) = \rho_{xx}(\tau) \quad . \quad . \quad . \quad . \quad . \quad . \quad (46)$$

This is not an easy problem, and as far as is known no rigorous explicit solution has been obtained. It is possible, however, to show that at least one important class of signals satisfies the first requirement. These are signals for which $\int_{-\infty}^{+\infty} xp(x, x_m)dx$ can be expressed in the form $P(x)Q(m)$, where $P(x)$ is a function of x only and $Q(m)$ is a function of m only.[10] This result was realized independently by the author, but was expressed rather differently in what is believed to be a more general and at the same time more useful form, namely that $\int_{-\infty}^{+\infty} xp(x|x_m)dx$ should be of the form $P(x)Q(m)$. Verification of either of these forms may be accomplished by substituting in either eqn. (41) or (43) as appropriate, and observing that $Q(m)$ may be moved outside the integral, thus leaving an integral which is independent of m.

An important form of $p(x, x_m)$ in the above category is the normal bivariate distribution

$$p(x, x_m) = \frac{1}{2\pi[1 - R_{xx}^2(\tau)]^{\frac{1}{2}}}\varepsilon^{[x^2+x_m^2-xx_mR_{xx}(\tau)]/2[1-R_{xx}^2(\tau)]} \quad . \quad . \quad . \quad . \quad (47)$$

The properties of this distribution for various types of distortion have been treated by several writers.[11,12] In the case of infinite clipping the correlation function of the output signal is

$$R_{yy}(\tau) = \frac{2}{\pi}\text{arc sin}\,[R_{xx}(\tau)] \quad . \quad . \quad . \quad (48)$$

An even more general class of signals has been defined.[10] It includes the previously mentioned class as a special case, but it is certainly not exclusive. Nevertheless, the formulation is so broad as to include, almost certainly, all cases of practical interest.

The second identity, of eqn. (46), (which inevitably requires that eqn. (45) be satisfied also) will be met if

$$\sigma_y^2\int_{-\infty}^{+\infty} xp(x|x_m)dx = \sigma_x^2\left[2\int_0^{\infty} p(x|x_m)dx - 1\right]$$
$$= S(x)T(m) \quad . \quad . \quad . \quad . \quad . \quad . \quad (49)$$

where $S(x)$ is a function of x only and $T(m)$ is a function of m only. (It should be noted, however, that this result is specific to the case of infinite clipping only.)

It is possible that there may be other probability distributions that do not meet the above requirements but still satisfy the relation of eqn. (46).

In conclusion, it is worth while emphasizing once again that although clipping may modify the spectrum of a partially constrained signal very considerably, the coherence coefficient still provides a reliable indication of the average common spectral energy.

(8) SUMMARY OF RESULTS

The signals treated in the paper may be divided into two principal categories, namely "totally random" and "partially constrained" signals. In both cases it is convenient to time-quantize the signal so that it consists of a train of adjacent rectangular pulses of varying amplitude [see Fig. 3(*a*)].

Two coherence coefficients are defined, namely μ and ν. The first, which is applicable to the case of straightforward clipping, is given by the maximum value of the cross-correlation function. The second is defined as the integral over all frequencies of the modulus of the cross-spectral-energy-density function.

The effect of infinite clipping on a totally random pulse train is considered. It is shown that

(*a*) The three correlation functions are related in the simple manner indicated by eqn. (23). They are all identical in shape (see Fig. 7), and this shape is a function only of the manner of quantization.

(*b*) The spectral-energy-density functions are therefore similarly related, and again identical in shape (see Fig. 8).

(*c*) On the assumption that the three restrictions of eqns. (1)–(3) are valid, the expression for the coherence coefficients is singularly simple, namely

$$\nu_{xy} = \mu_{xy} = \frac{2}{\sigma_x}\int_0^{+\infty} xp(x)dx \quad . \quad . \quad . \quad . \quad (50)$$

where σ_x is the standard deviation of the probability distribution $p(x)$ defining the range of amplitudes x of the original signal.

(*d*) When $p(x)$ is Gaussian in form [see eqn. (4) and Fig. 4], μ_{xy} has the value 0·798. For an exponential distribution [see eqn. (5) and Fig. 4] μ_{xy} has the value 0·707. It is interesting to note that in both cases the coefficients are relatively large.

The effect of subsequent differentiation is considered. Expressions for the correlation functions and also the spectral-energy-density functions are deduced, and are illustrated in Figs. 9, 10 and 11. The only simple relation between these functions is that the spectral-energy-density function of the output signal is identical in shape to the real part of the cross-spectral-energy-density function. It is shown that the first coherence coefficient μ is reduced by a factor of $\sqrt{2}$ as a result of subsequent differentiation. By contrast, the second coherence coefficient is reduced by only 10%.

The third restriction on the form of $p(x)$ expressed in eqn. (3) is relaxed. The only change is that the coherence coefficients are modified in the simple manner indicated in eqns. (36) and (37). Whereas asymmetry increases the coherence for simple clipping, it decreases the coherence when the clipped signal is subject to subsequent differentiation.

The mathematical techniques previously described are applied to the case of partially constrained signals. It is shown that in the case of simple clipping the coherence coefficient is defined as before and is independent of the nature of the statistical constraints relating the values of successive signal amplitudes. A quantized technique is indicated for evaluating the spectral-energy-density functions and correlation functions of clipped, partially constrained signals. It is shown that there is no simple criterion for determining the conditions under which the signal is least likely to be mutilated by the process of clipping. However, certain special cases can be formulated, amongst which is notably the normal bivariate distribution.

(9) CONCLUSIONS

(9.1) Speech Clipping

It is important to examine to what extent the results obtained may be applied to the case of clipped speech.

In the first place it is evident that speech falls into the category of partially constrained signals. However, this by itself does not constitute a limitation, since it has been shown in Section 7.1 that the first coherence coefficient is unaffected by the character of the statistical constraints between successive signal values.

Secondly, it is clear that the results only have significance in respect of very long samples. In other words, $p_1(x)$ and $p_2(x)$ in eqns. (4) and (5) are the respective averages over all the possible unvoiced and voiced sounds. With this formulation nothing can be deduced about the intelligibility of individual sounds, except of course that large deviations below the average must be relatively infrequent.

The principal difficulty is the unknown relationship betwe the coherence coefficients and intelligibility. In the case clipped speech the theoretical values of coherence coefficie are 80% for unvoiced sounds (e.g. fricatives) and 71% for voic sounds (e.g. vowels, semi-vowels, etc.). The measured ar culation scores for single-syllable words (voiced and unvoic sounds mixed) rises from about 70% to a maximum of 90% the test proceeds, after which there is a falling off owing fatigue.[1] Further investigation[3] has shown that the vowel soun suffer more than consonants; this is in general accord with t relative values of coherence coefficient for unvoiced and voic sounds. The effect of subsequent differentiation is to redu the first coherence coefficient by a factor of $\sqrt{2}$, and the seco coherence coefficient by only 10%. Articulation tests[1] show no significant change due to subsequent differentiation, whi would seem to justify the use of the second coherence coefficie in this case.

On the basis of the above figures it is tempting to assume on a *post hoc ergo propter hoc* basis—a denite formal relationsh between coherence coefficient and intelligibility. This wou indeed be dangerous, but on very general grounds it does see likely that the coherence coefficient must provide an approxima first-order indication of intelligibility. A more accurate criteri would be based on the correlation function as a whole, and al the higher-order product moments.

The margin between the coherence coefficient and the ar culation score may in part be due to the marked asymmetry the average speech waveform, which must be taken into accou in the numerical calculation of the coherence coefficient. Ho ever, the greater part of the margin must be attributed to t ability of the ear to utilize the inherent redundancy in the wav form to combat the effect of signal distortion.

With regard to the relationship between the auto-correlati functions before and after clipping, it is interesting to note th some preliminary experiments with a computer described in companion paper[8] have shown a marked resemblance for tho sounds so far tested.

(9.2) Further Pertinence of the Results

Clipping may be utilized as a form of coding, i.e. it resembles pulse-code-modulation system in which the signal is represent by a single binary digit. The percentage harmonic distorti due to sampling is then $\sqrt{[2(1-\mu)]}$, where μ is the coheren coefficient as defined in Section 3.3.

The mean-square error and the correlation coefficient a related in the following manner:

$$\overline{(x-y)^2} = \overline{x^2} + \overline{y^2} - 2\,\overline{yx}$$

Hence it is possible that a few of the results may be of use in t calculation of the response of highly non-linear servo mechanis to random inputs. In particular, the technique of quantiz tion outlined in Section 7 may assist in simplifying speci computations.

(10) ACKNOWLEDGMENTS

The investigation described in the paper was stimulated by temporary association with a programme of research at t Imperial College of Science and Technology into the facto affecting the intelligibility of speech, and was undertaken on t author's own initiative. The author would like to thank Dr. E. Cherry, and Messrs. A. J. Fourcin and J. N. Holmes for ma stimulating discussions and much helpful advice in the prepar tion of the paper. The solution to an integral in Section 5 was suggested by Mr. L. Lewin. The author is grateful Standard Telecommunication Laboratories for permission publish the paper.

(11) REFERENCES

) LICKLIDER, J. C. R., and POLLACK, I.: "Effect of Differentiation Integration and Infinite Clipping upon the Intelligibility of Speech," *Journal of the Acoustical Society of America*, 1948, **20**, p. 42.

) LICKLIDER, J. C. R.: "The Intelligibility of Amplitude-Dichotomised Time-Quantised Speech Waves," *ibid.*, 1950, **22**, p. 820.

) Unpublished work at the Imperial College of Science and Technology, London.

) GOLDMAN, S.: "Frequency Analysis, Modulation and Noise" (McGraw-Hill, New York, 1948), p. 333.

) JAMES, H. M., NICHOLS, N. B., and PHILLIPS, R. S.: "Theory of Servo-Mechanisms," *M.I.T. Radiation Laboratory Series No.* 25 (McGraw-Hill, New York, 1947), Chapter 6.

) WIENER, N.: "The Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications" (Wiley, New York, 1949).

(7) DUKES, J. M. C.: "A Simple Correlation Function Analogue Computer," Thesis submitted for the Diploma of the Imperial College of Science and Technology, January, 1951.

(8) HOLMES, J. N., and DUKES, J. M. C.: "A Speech-Waveform Correlator with Magnetic-Tape Delay and Electronic Multiplication," *Proceedings I.E.E.*, Paper No. 1639 R, July, 1954 (**101**, Part III, p. 225).

(9) DAVENPORT, W. B.: "A Study of Speech Probability Distributions," *Journal of the Acoustical Society of America*, 1952, **24**, p. 390.

(10) LUCE, R. D.: "Amplitude Distorted Signals," *Quarterly Progress Report of the Research Laboratory of Electronics, M.I.T.*, April, 1953, p. 37.

(11) BUSSGANG, J. J.: "Cross-Correlation Functions of Amplitude-Distorted Gaussian Signals," Research Laboratory of Electronics, M.I.T., Report No. 216, March, 1952.

(12) VAN VLECK, J. H.: "The Spectrum of Clipped Noise," Radiation Research Laboratory, Harvard, Report No. 51, July, 1943.

621.372.2.029.65

Monograph No. 112
Dec. 19

THE GENERATION OF MILLIMETRE WAVES

By J. L. FARRANDS, Ph.D., B.Sc.

(*The paper was first received* 15*th March, and in revised form* 9*th August*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in December*, 1954.)

SUMMARY

Problems in the generation of millimetre waves are considered, and further study of sparked generators is found desirable. A contribution is made to the simple theory of such generators. Experiments made to verify the theory are not wholly successful, but their evidence, together with that of other workers, strongly suggests that this simple theory is largely correct. Experiments at variance with the present results are examined critically. It is concluded that useful results can only be obtained with sparks in oil.

LIST OF SYMBOLS

μ = Permeability in boundary-value problem.
ϵ = Permittivity in boundary-value problems.
l = Characteristic length.
τ = Characteristic time (e.g. period of the field).
C_1, C_2 = Parameters required to be constant for similarity of boundary-value problems.
σ = Conductivity in boundary-value problem.
ω = Angular frequency.
a = Radius of sphere.
t = Time.
$E(t)$ = Radiated field at time t.
α = Damping coefficient of natural oscillations of sphere.
ω_0 = Natural angular frequency of sphere.
$G(\omega)$ = Amplitude of component having frequency ω in the spectrum of $E(t)$.
$G^*(\omega)$ = Complex conjugate of $G(\omega)$.
$\Delta\omega$ = Element of frequency selected by a spectrometer.
$P(\omega)$ = Power radiated in the element $\Delta\omega$.
m = Resolving power of the spectrometer.
n = A parameter given by $\alpha = n\omega_0$.
p = Dipole moment of a metallic sphere in an electric field E.
W = Potential energy of the sphere in a static field.
s = Number of discharges per second.
η = Efficiency of energy radiation.
V_1, V_2 = Output and input amplitudes of pulses to amplifier.
B = Amplifier bandwidth.
γ = Crystal parameter—volts video output per watt of r.f. input.

(1) DIFFICULTIES WITH COHERENT OSCILLATORS

The region of the electromagnetic spectrum between the wavelengths of about 200 microns and a few millimetres remains relatively unexplored, even after half a century of radio research. This region is at present of special interest to physicists, biophysicists, engineers and chemists, because of the many problems of physical structure and instrumentation whose solution would be greatly facilitated by suitable sources and detectors for electromagnetic waves in this part of the spectrum.

The recent extension of the infra-red techniques to 700 microns by McCubbin[1] is an important contribution to the subject, but the extension of the radio spectrum has met with considerable difficulties. Attention has been drawn to the problems Pearce[2] and Klumb,[3] and it is not intended to restate their vie here; however, it is desirable at this stage to examine some the suggested methods and to examine critically some of work that has been reported.

Necessary properties are, of course, that source and detec should have sufficient power and sensitivity to allow for refl tions and power losses in the apparatus and material bei examined. It is further desirable (e.g. in the examination fine structures) that the radiation used should have as narrow bandwidth as possible. This latter requirement can be met the application of filtering techniques (e.g. spectrometers) wideband sources, but for the sake of efficiency it would better if the source itself were a so-called "monochromati source such as, for example, a klystron, or to a lesser degr a magnetron. Consideration of the probable sensitivities thermal detectors (Golay cell, thermocouple, etc.) and cryst in this part of the spectrum leads to the expectation that t minimum useful resolved power is of the order of a microwa

The scaling of conventional electronic devices is impeded the nature of materials available. Stratton[4] has shown that t boundary-value problems are similar if the following relatio hold:

$$\mu\epsilon\left(\frac{l}{\tau}\right)^2 = C_1 \quad . \quad . \quad . \quad . \quad . \quad .$$

$$\mu\sigma\frac{l^2}{\tau} = C_2 \quad . \quad . \quad . \quad . \quad . \quad .$$

In constructing a shorter-wavelength scale model of an osc lator, eqn. (1) is easily maintained by making l proportional the wavelength. Such a transformation will not satisfy eqn. (however, unless the quantity $\mu\sigma$ is increased in the ratio of t wavelength. With present-day materials this is impossible unl superconductors are employed; indeed, far from $\mu\sigma$ increasi at decreasing wavelengths, it decreases by the loss in "effect conductivity" (i.e. high-frequency measured value) owing to t deterioration of surface finish relative to the wavelength. T overall result of this is, of course, an increase in the losses of t inert circuit of the scaled oscillator, so that the oscillations beco more and more feeble at shorter wavelengths until they eventua cease.

Apart from the somewhat fundamental difficulty mention above, there are those mentioned by Pearce.[2] Briefly these a first, that the electron-beam current varies as the square of t wavelength if present current densities are maintained, a secondly, that the dissipation of heat varies linearly with t wavelength for conduction and as the square of the wavelen for radiation. The deterioration of heat-loss efficiency leads higher and higher temperatures of components at shorter wa lengths until eventually the melting points of present materi are reached.

These considerations lead to the unhappy conclusion that t scaling of conventional oscillators such as klystrons is unlik to go very far towards closing the gap in the spectrum. It

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Farrands is at the Defence Standards Laboratory, Australia, and was formerly at the Imperial College of Science and Technology.

obable that the limit is reached at wavelengths of a few illimetres.

For certain applications, e.g. microwave spectroscopy of gases, has been possible to retain the advantages of coherent sources frequency multiplication[5,6] in non-linear elements or by the rect use of harmonics.[7] These techniques are of limited plication because of the very low powers involved, but they pear to be capable of reaching wavelengths of between one d two millimetres.

It would appear that progress in this spectral region with herent oscillators must await some really novel generators or storters. In the meantime, it is necessary to reconsider the acticability of incoherent oscillators for investigations.

(2) SPARKED RESONATORS

The oldest and most successful generator of short electromagnetic waves is the sparked resonator. One or more resonators arbitrary shape are placed between a pair of electrodes and arks are allowed to pass between them. The sparks are companied by damped trains of waves of random phase. he prototype of this is the Righi–Hetz dipole, and oscillators this kind were used very successfully by Nicholls and ear[9,10] to produce waves about 2mm long as early as 1923. Many other workers have employed sparked resonators.[11–16] onsiderable ingenuity has been shown by these workers in eveloping resonators which can suffer frequent sparking with minimum of erosion. The resonators of Nicholls and Tear ere short thick linear dipoles sealed in glass; Arkadiewa[12,13] sed a mixture of metal filings in oil (the so-called "mass diator"), while Lewitsky,[11] and later Rohrbaugh,[15] used a eometrical array of resonators on a glass plate (the so-called matrix radiator"); Daunt used lead shot in an oil stream—a ariation of the mass radiator.

In spite of the considerable effort expended over nearly half a entury, it must be admitted that the phenomena of these sparked scillators are only vaguely understood. New workers in the eld find very little explicit and unequivocal information, and any questions arise immediately. The first is whether all these chniques are the same in principle. The simplest picture of ny of them is to imagine an inhomogeneous distribution of harges induced on the resonator by the field between the electodes; the passage of the spark causes the binding field to ollapse, and if the rate of collapse is sufficiently great, there will e an oscillatory redistribution of charge with radiation. Alternately the resonator may be regarded as a filter which separates ut and readily radiates the desired components of frequency ntained in the current pulse of the spark.

Such a model is not supported by the experiments of Kelliher nd Walton,[17] which suggest that a mechanism exists which nds to maintain the oscillations of the resonator. Further uestions arise, such as whether regular and predictable spectra an be obtained with regular resonators; whether the resonators multiple systems, such as the mass radiator and matrix diator, oscillate singly or in pairs and groups as suggested by rkadiewa;[12] the effect of the surrounding medium and whether arks in air can be used as reported by Rohrbaugh or whether l must be used as in the methods of all other workers; whether fficient power can be generated for useful absorption measureents, and whether the power can be calculated.

Considerable technical advantages would be obtained by using arks in air. Calculations are simplified because no unknown rmittivities are introduced and experiments are simpler. arks through oil generate dirty carbon suspensions and emical impurities, very small amounts of which greatly modify e conditions of experiment. Apart from this, the explosive nature of a spark in oil precludes the possibility of rigidly mounting resonators while introducing a minimum of perturbation to their characteristic frequencies.

There is not, so far, sufficient experimental evidence to answer all the questions precisely, particularly since the reports in the literature are often fragmentary as to method and frequently contradictory. An effort is made in the following Sections to obtain the maximum amount of information from some of the reports. As is well known, boundary-value problems involving cylinders and rods are not easy to solve, and so it will be necessary to obtain most of the information from experiments using spheres, since these provide the opportunity to compare the experimental results with theories which may be derived.

(3) SPECTRUM RADIATED BY A SPHERE

It has long been known[18,19] that the natural oscillations of a perfectly conducting sphere in its lowest mode are described by the function

$$\exp(-i\omega t) = \exp\left(-\frac{0{\cdot}5}{a} \mp i\frac{0{\cdot}86}{a}\right)\sqrt{(\epsilon\mu)}t \quad . \quad (3)$$

where ϵ and μ are the permittivity and permeability of the medium in which the sphere is embedded. The damping term in the expression arises from the radiation which is so efficient that the introduction of the finite conductivity of the sphere into the expression has little effect.

If the radiation from a sparked sphere were entirely due to the redistribution of charges freed by the removal of the field, the radiated field might be expected to have the same form, namely

$$E(t) = 0, \quad t < 0$$

$$E(t) = \epsilon^{-\alpha t}\sin\omega_0 t \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where

$$\alpha = \frac{0{\cdot}5}{0{\cdot}86}\omega_0$$

The Fourier transform of this wave is

$$G(\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} E(t)\epsilon^{-i\omega t}dt \quad . \quad . \quad . \quad . \quad (5)$$

and the power product is

$$G(\omega)G^*(\omega) = \frac{\omega_0^2}{4\pi^2[(\alpha^2+\omega_0^2-\omega^2)^2+4\alpha^2\omega^2]} \quad . \quad (6)$$

So that the power radiated in the frequency range

$$\omega - \frac{\Delta\omega}{2} < \omega < \omega + \frac{\Delta\omega}{2}$$

is given by

$$P(\omega) = \frac{\omega_0^2\Delta\omega}{4\pi^2[(\alpha^2+\omega_0^2-\omega^2)^2+4\alpha^2\omega^2]} \quad . \quad . \quad (7)$$

For gratings and spectrometers generally, it can be assumed—at least over a small range—that the resolving power is a constant, i.e.

$$\Delta\omega/\omega = m$$

so that eqn. (7) becomes

$$P(\omega) = \frac{\omega_0^2}{4\pi^2[(\alpha^2+\omega_0^2-\omega^2)^2+4\alpha^2\omega^2]}\omega m \quad . \quad (7a)$$

and the apparent-power distribution curve becomes

$$\frac{P(\omega)}{P(\omega_0)} = \frac{\alpha^2(\alpha^2-4\omega_0^2)\omega}{[(\alpha^2+\omega_0^2-\omega^2)^2+4\alpha^2\omega^2]\omega_0} \quad . \quad . \quad (8)$$

Eqn. (8) can be rearranged and normalized by the substitution $\alpha = n\omega_0$ to the form

$$\frac{P(\omega)}{P(\omega_0)} = \frac{n^2(n^2+4)\left(\frac{\omega}{\omega_0}\right)}{\left[n^2+1-\left(\frac{\omega}{\omega_0}\right)^2\right]^2 + 4n^2\left(\frac{\omega}{\omega_0}\right)^2} \quad . \quad . \quad (9)$$

This function, with the value of n appropriate to spheres, is shown in Fig. 1. The points in triangles are obtained from the experimental results of Daunt *et al.*[16]

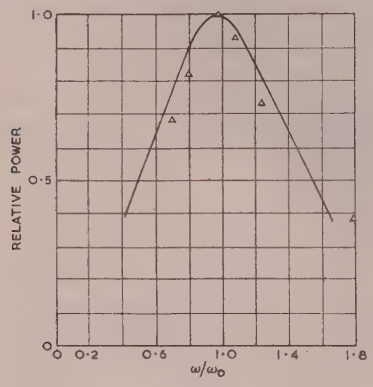


Fig. 1.—Theoretical curve of spectral distribution of energy.
△ Normalized values from Daunt *et al.*

The experimental points appear to suggest a spectrum slightly narrower than the theoretical one. It should be remembered, however, that the detector used was almost certainly not uniform in its response, probably falling off in sensitivity at shorter wavelengths, and that the grating efficiency probably declines at long wavelengths (i.e. large angles of diffraction). The high point at $\omega/\omega_0 = 1{\cdot}8$ can probably be ascribed to harmonic radiation.

On the whole, the agreement between the experimental results and the predictions of the simple theory are surprisingly good. The agreement strongly suggests that if there is an oscillation-maintaining effect in the spark it has negligible effect. The point may be raised as to whether the effect observed by Kelliher and Walton[17] was due to filtering, possibly through some resonance in the detector assembly.

Information of interest in another field of physics appears to be derivable from this result. It is implicit in eqn. (4) that the binding field is removed instantaneously. In practice, of course, this is not so, as the field decays in the time of build-up of the spark. The agreement found above suggests then that the time of breakdown is not greater than a quarter of a period of the waves used by Daunt *et al.*,[16] i.e. the breakdown time in the oil used by them may well have been less than 4×10^{-12} sec.

(4) THE AVAILABLE POWER

The success of the simple picture in predicting the frequency radiated by the sparked spheres suggests that it may also be used to calculate the available power.

The charge distribution on a metallic sphere between spherical electrodes presents a problem of considerable complexity which it is not worth while to solve accurately in the present case. A sufficient approximation is obtained by considering the sphere to be immersed in a homogeneous field. It is known that a field E through a medium with permittivity ϵ will induce in a conducting sphere of radius a a dipole moment p, where

$$p = 4\pi\epsilon E a^3 \text{ (in M.K.S. units)} \quad . \quad . \quad . \quad (10)$$

The potential energy of this dipole is

$$W = pE \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1$$

and presumably this is the maximum amount of energy that ca be radiated in a single discharge. The appropriate value of is the breakdown field strength of the medium in which th sphere is immersed. If there are s discharges per second th average power radiated is

$$P = pE\eta s$$
$$= 4\pi\epsilon\eta s E^2 \quad . \quad . \quad . \quad . \quad . \quad (1$$

where η is the efficiency, i.e. the ratio of energy radiated to th original energy of the sphere. (The absence of significa damping from the radiated spectrum suggests that η is ve large—nearly 100%.)

It was stated earlier that it is desirable to carry out experimen with sparks in air, so that it is now worth while inserting son possible values in eqn. (11).

Taking $E = 3 \times 10^5$ volts/m

$$\epsilon = \frac{1}{36\pi} \times 10^{-9} \text{ farad/m}$$

$$a = 2{\cdot}5 \times 10^{-4} \text{ metre}$$

$$\eta = 1$$

Eqn. (11) yields

$$W = 16 \times 10^{-11} \text{ joule per spark per sphere}$$

and from eqn. (12),

$$P = 10^{-7} \text{ watt, for 3 spheres at 200 c/s.}$$

This is a very small signal, but it may be in considerable err for several reasons. The first is, of course, that the breakdow strength of air may greatly exceed 3×10^5 volts/m if sufficient sharp rising pulses are used; secondly it may not be correct assume that η is nearly unity, as an increase in damping may obscured by the experimental effects referred to in Section 3.

Not all the energy radiated can be used for measurement If an optical spectrometer is used, there are the apertures radiating and detecting mirrors and the relative inefficiencies diffraction gratings. (It would be unwise to use an echel grating of high efficiency with such a wide spectrum, since erro of interpretation would arise through the overlapping of highe order diffraction maxima.) It is estimated that the signal du to all frequencies in a single order would be reduced to abo 5% of its original value through these causes alone.

Dispersion of the spectrum to give a usefully narrow line by grating of, say, 100 lines (i.e. $\Delta\omega/\omega = 1/100$) leads to a resolve power at centre frequency, ω_0, of

$$\frac{P(\omega_0)\Delta\omega_0}{\int_0^\infty P(\omega)d\omega} \simeq \frac{1}{50} \quad . \quad . \quad . \quad . \quad . \quad (1$$

These corrections lead to a resolved power in the experime proposed of 10^{-10} watt. Even this ignores the possibility smaller signals due to inadequate deionization of the gap 200 c/s, and further, the difficulty of focusing all the energy the detector since the detector size is not fixed.

(5) DETECTORS

A practical difficulty in working in this part of the spectrum that the properties of detectors are unknown and remain

ıtil known amounts of power can be generated and properly tenuated. In the absence of single-frequency oscillators, it is, course, impossible to use the sensitive techniques of super-terodyne detection. Consequently, it is necessary to fall back ı either thermal detectors or crystal rectification.

Thermal detectors, such as the Golay cell,[23] have a sensitivity about 5×10^{-11} watt in the infra-red region, but it is not ıown whether this sensitivity is maintained at wavelengths of ɔout a millimetre. On the whole, it is suspected that it is not, ɛcause of the effect of aperture size and resonant effects in the ıvity. A characteristic of this type of detector is its long time-ɔnstant (about $\frac{1}{10}$sec), which not only makes for tedious ɛperiment but is a real disadvantage with sparked oscillators hose life is very short.

An interesting extension of the normal use of crystal detectors ɔr this type of problem has been made by Dickey[20] using a ɛsult of Montgomery.[21] It is found that if a pulse of duration τ nd amplitude V_1 is incident on a filter of bandwidth B, the output ulse is of amplitude V_2, where

$$V_2 = \tfrac{3}{2}B\tau A_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

B is much less than the spectral width of the incident pulse.

The output pulse of a crystal is proportional to the input ower,

e. $$V_1 = \gamma P$$

here γ is a constant of proportionality in volts per watt, and is the input power, so that eqn. (14) may be written

$$V_2 = \tfrac{3}{2}\tau B\gamma P$$

r $$V_2 = \tfrac{3}{2}B\gamma W \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

here $W = P\tau$ is the pulse energy.

If the spreading resistance of the crystal is small compared with ıe non-linear barrier resistance, and if the frequency is in excess f a few thousand megacycles per second, γ is expected to vary s the square of the frequency.[20, 22] Extrapolation of charac-ɛristic properties to wavelengths near a millimetre suggests that will be approximately equal to 100 mV/watt.

The use of such a crystal with a single-sphere pulse, as cal-ulated in the previous Section, would lead to a signal of very early 100 μV for the direct signal having a bandwidth of 4 Mc/s. ıfter frequency separation by the grating, this signal would be ɛduced to about 0·1 μV.

While it must be realized that the results given are, at best, pproximations, it will be appreciated that the prospect of btaining useful amounts of power in this part of the spectrum not great. However, the calculations are based on the most mple hypothesis, and although this has received partial con-rmation through the spectrum experiments, it has to be faced ıat experiments have been reported using sparks in air. Con-ɛquently the calculation above cannot be accepted until further xperimental work has been done.

6) EXPERIMENTS AT THE IMPERIAL COLLEGE OF SCIENCE AND TECHNOLOGY

Attempts to verify the predictions of the simple theory of the parked resonators in air with a matrix radiator were not wholly ıccessful. Hard steel balls tuned to 1·8 mm were mounted on a lass plate. (The resonators of Lewitsky were fixed to a glass late with Canada balsam, and the irregular chromium particles f Cooley and Rohrbaugh were partially embedded in the glass.) n the present case very small indentations were etched in the lass; the indentations were filled with a thermosetting plastic dhesive to bind the metal spheres without immersing them in ıe adhesive itself.

It was found that these assemblies of resonators had a limited life owing either to the erosion of the spheres or, more rarely, to failure of the adhesive. It was also found that, depending on the current allowed to pass in the spark, the life was limited by a change in the surface of the glass at high spark rates. This change, which was not permanent, appeared to be due to heating of the glass surface providing a conducting path, so that the discharge no longer took place through the air but over the surface. The attainable voltage was considerably reduced, and the glass frequently cracked under these conditions.

The voltage source was a pulse modulator similar to those used for magnetrons. The pulse length was either 2·5 or 5 microsec, and the pulse repetition frequency was variable. The source and detector were located at the foci of 20 in aperture front-silvered mirrors. Gratings were made from tin-foil mounted on expanded Bakelite sheet.

Attempts made with a Golay cell[23] were wholly unsuccessful, while experiments with a crystal detector were not much better, but some results were obtained. The signal from the crystal was amplified in a 4 Mc/s bandwidth amplifier at 60 Mc/s and displayed on a high-speed oscillograph. This oscillograph display was a considerable advantage, as it enabled the signal due to the spheres themselves to be separated from other unshielded signals accompanying the discharge.

Signals obtained with the grating in the zero-order (specular reflection) setting were of the order of 200 μV input to the amplifier. No signals could be detected in the dispersed spectrum from the grating. Attempts made to discover the frequency of the radiation by using a Boltzmann[24] interferometer were unsuccessful owing to the extreme fluctuations in signal amplitude; indeed it was impossible to learn by this means anything more than that the wavelength was considerably less than 1 cm.

While a measurement of wavelength would be definite proof of the nature of the radiation, there is nevertheless no doubt that the radiation emanated from the spheres themselves. This is shown by experiments with shielding and by the fact that the signal disappeared if the spheres were removed from the spark. (That this was not due to a change in the conditions of the spark was shown by readjusting the gap between electrodes to the same dimensions as that formerly existing between the spheres.)

It was found that the signal amplitude increased with the number of spheres up to about three; however, this increase appeared to be less than linear with the number of spheres. This is to be expected, of course, because of the difficulty of focusing an extended line-source on the crystal. On the other hand, it suggests that the effect of cascaded breakdown of the gaps is probably less than might have been expected.

These experiments appear to confirm, at least in part, the consequences of the hypothesis. The position is not wholly clear, of course, since considerable extrapolation was used and many assumptions were made. It is extremely unlikely, in fact, that the breakdown strength of air is as low under pulse conditions as has been assumed, so that either the radiation process is less efficient than assumed or the detector is less sensitive. Both may be true. On the whole, however, the agreement is felt to be good.

In spite of this conclusion, the fact remains that another group of workers has made successful experiments with a dispersed spectrum from sparks in air, and since this represents, at first sight, a contradiction of the theory, it is necessary to examine their results in more detail.

(7) EXPERIMENTS OF COOLEY AND ROHRBAUGH, *ET AL.*

The transmission curves shown in Fig. 2 were obtained at New York using an array of irregular chromium particles embedded in a glass plate. Sparks were passed through the air

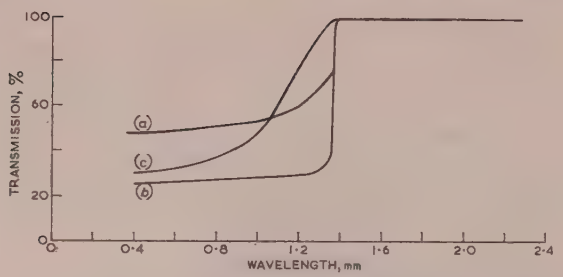


Fig. 2.—Transmission curves for polystyrene obtained by Cooley and Rohrbaugh, *et al.*

(*a*) ¼in sheet.
(*b*) Two ¼in sheets together.
(*c*) ⅛in sheet from another manufacturer.
This diagram is reproduced from *Journal of Applied Physics*, 1948, **19**, p. 1093.

between the particles. The spectrum was dispersed by a grating. It is understood that a thermal detector was used. The transmission measurements were made on flat sheets ¼in thick.

Several features of these curves should be noted. The first is the quite unexpected very sharp edge in the Amphenol samples. The very high transmission at wavelengths greater than 1·3mm is surprising in view of the necessary reflection losses. A final observation is the absence of interference maxima and minima, particularly in the transparent region of the curves; it is unlikely that these effects would be beyond the resolution of the spectrometer employed. (Transmission is claimed to be measured to 5%, so that such effects should be readily observable in a material with the permittivity of polystyrene.)

McCubbin[25] has measured the properties of polystyrene with his long-wave infra-red spectrometer, and these results are presented in Fig. 3, together with those of Cooley and Rohrbaugh,

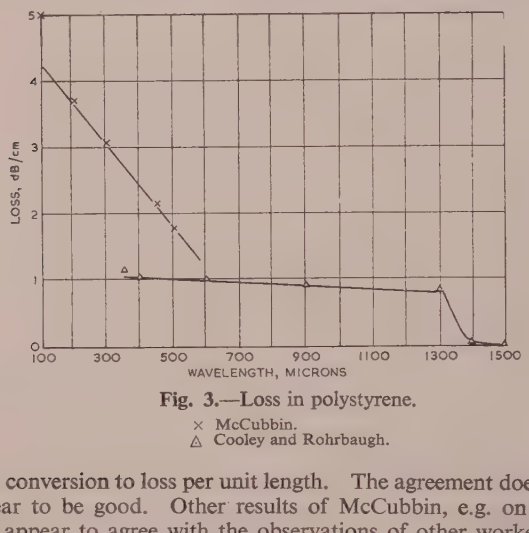


Fig. 3.—Loss in polystyrene.

× McCubbin.
△ Cooley and Rohrbaugh.

after conversion to loss per unit length. The agreement does not appear to be good. Other results of McCubbin, e.g. on rock salt, appear to agree with the observations of other workers in the infra-red and in the radio spectrum.

It would seem from these remarks that the measurements made by Cooley and Rohrbaugh can only be accepted with considerable caution.

Presumably those workers could have achieved greater radiated powers than in the present work by the use of faster rising voltages, although there is probably not a great gain since even the fastest rising pulses will be considerably rounded by the capacitance of the spheres. Furthermore, any increase in energy due to this cause would be considerably offset by the use of irregular particles, as irregularities would decrease the brea down voltage and also distribute the energy over a wid spectrum.

(8) CONCLUSION

It is unfortunate that neither the experiments at the Imper College of Science and Technology nor those carried out overs are wholly conclusive, but it is felt that the weight of evidence in favour of the simplest hypothesis. Since one of the cons quences of this hypothesis is that sufficient power for measur ment purposes is unlikely to be obtained using sparks in a there seems to be little point in pursuing this line of research.

There can be no doubt that the sparking of resonators in is successful, as shown by Nicholls and Tear and later by Dau *et al.* There is scope for quantitative work on the ener radiated by these systems, and for research into the speed breakdown of the oil. A final conclusion is that a considerab improvement in the performance of mass radiators could obtained by using regular spherical radiators instead of t irregular metal filings used by Glagolewa-Arkadiewa.

(9) ACKNOWLEDGMENTS

The author wishes to express his thanks to Prof. Wil Jackson, and Mr. J. Brown, of the Imperial College Science and Technology, for their assistance in this work an in the preparation of the paper. He also wishes to than Imperial Chemical Industries, Ltd., for a grant for the purcha of equipment, and the Public Service Board of Australia for the generous scholarship.

(10) REFERENCES

(1) McGubbin, Jr., T. K., and Sinton, W. M.: "Far Infra-re Spectroscopy from 100 to 700 microns," *Journal of t Optical Society of America*, 1952, **42**, p. 113.
(2) Pierce, J. R.: "Millimetre Waves," *Physics Today*, Nov 1950.
(3) Klumb, H.: "Intensitätsfragen im Gebiet sehr kurz elektrischer Wellen," *Zeitschrift für Physik*, 1940, **11** p. 321.
(4) Stratton, J. A.: "Electromagnetic Theory" (McGrav Hill, New York, 1941), p. 558.
(5) Gillam, C. R., Johnston, C. M., and Gordy, W.: "Micr wave Spectroscopy in the Region from 2–5mm," *Physic Review*, 1950, **78**, p. 140.
(6) Smith, A. G., Gordy, W., Simmons, J. W., and Smit W. V.: "Microwave Spectroscopy in the Region 3–5mm," *ibid.*, 1949, **75**, p. 260.
(7) Loubser, J. H. P., and Townes, C. H.: "Spectroscop between 1·5 and 2·0mm using Magnetron Harmonics, *ibid.*, 1949, **76**, p. 178.
(8) Lebedew, P.: "Über die Doppelbrechnung der Strahle electrischer Kraft," *Annalen der Physik und Chemie*, 189 **56**, p. 1.
(9) Nicholls, E. F., and Tear, J. D.: "Joining the Infra-re and Electric Wave Spectra," *Astrophysical Journal*, 192 **61**, p. 17.
(10) Nicholls, E. F., and Tear, J. D.: "Short Electric Waves, *Physical Review*, 1923, **21**, p. 587.
(11) Lewitsky, M. A.: "Elektrische Wellen im Gebiete de äussern Ultrarote," *Physikalische Zeitschrift*, 1926, **2** p. 177.
(12) Glagolewa-Arkadiewa, A.: "Short Electromagnet Waves of Wavelength up to 82μ," *Nature*, 1924, **11** p. 610.
(13) Glagolewa-Arkadiewa, A.: "On the Theory of the Ma Radiator," *Comptes Rendus de l'Académie des Scienc de l'U.R.S.S.*, 1941, **32**, No. 8, p. 540.

) COOLEY, J. P., and ROHRBAUGH, J. H.: "The Production of Extremely Short Electromagnetic Waves," *Physical Review*, 1945, **67**, p. 296.
) WANTUCH, E., COOLEY, J. P., ROHRBAUGH, J. H., GREIG, J. H., and SHMOYS, J.: "Transmission Properties of Materials in the Millimetre Wave Region," *Journal of Applied Physics*, 1948, **19**, p. 1092.
ROHRBAUGH, J. H.: "A Study of the Generation and Detection of Electromagnetic Waves in the Millimetre Region," Final Report Contract No. W28–099 ac-171, New York University, 1948.
) DAUNT, J. G., JACKSON, D. A., HULL, R. A., and HURST, C.: "Experiments on the Production, Detection and Measurement of Millimetre Radiation and its Absorption by the Atmosphere," Central Valve Development Committee, Ministry of Supply, Report CL Misc. 44, May, 1945.
7) KELLIHER, M. C., and WALTON, E. T. T.: "Micro Electromagnetic Waves," *Wireless Engineer*, 1946, **23**, p. 46.
8) THOMSON, J. J.: "Recent Researches in Electromagnetism" (Oxford University Press, 1893), p. 370.
(19) STRATTON, J. A.: "Electromagnetic Theory" (McGraw-Hill, New York, 1941), p. 558.
(20) DICKEY, JR., F. R.: "The Production of Millimetre Waves by Spark Discharges," Technical Report No. 123, Cruft Laboratory, Harvard University, 1951.
(21) MONTGOMERY, C. G.: "Technique of Microwave Measurements," *M.I.T. Radiation Laboratory Series* (McGraw-Hill, New York, 1948), p. 188.
(22) TORREY, H. C., and WHITMER, C. A.: "Crystal Rectifiers," *M.I.T. Radiation Laboratory Series* (McGraw-Hill, New York, 1948), p. 337.
(23) GOLAY, M. G. D.: "A Pneumatic Infra-red Detector," *Review of Scientific Instruments*, 1947, **18**, p. 347.
(24) FARRANDS, J. L., and BROWN, J.: "Boltzmann Interferometer-Interpreting Interferograms," *Wireless Engineer*, 1954, **31**, p. 81.
(25) MCCUBBIN, T. K., JR., and SINTON, W. M.: *Journal of the Optical Society of America*, 1950, **40**, p. 537.
(26) POTOK, M. N. H.: "A Critical Review of Researches into Millimetric-Wave Sparked Generators," *Journal of the British Institution of Radio Engineers*, 1953, **13**, p. 490.

621.396.674.3

Monograph No. 11
Dec. 19

THE RADIATION OF A HERTZIAN DIPOLE OVER A COATED CONDUCTOR

By D. B. BRICK, Ph.D., S.M., A.B.

(*The paper was first received 8th January, and in revised form 2nd July*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in December*, 1954.)

SUMMARY

The idealized problems of an infinitesimal Hertzian dipole in and over a perfect dielectric coating a perfect conductor and an Abraham dipole lying on the conductor are treated. Unintegrated forms of the Hertz potentials are obtained both for electric and magnetic dipoles.

Integrated far-zone forms of the potentials and fields are obtained for electric dipoles by means of asymptotic integrations. Far-zone radiation patterns are given in order to indicate the distortions of the fields and the magnitudes of the residue waves caused by the dielectric coatings. These show that such layers cause a very large increase in the relative field strength in the direction along the surface.

It is proved that the power radiated by the dipole may be divided into two independent quantities—the power fed to radiation-type and that fed to surface or guided-type fields. For certain cases numerical results are given for the total power radiated and the relative powers fed to the two types of fields.

Formulae are derived and illustrated with numerical examples of the radiation resistances of the dipoles and the attenuation coefficients of the surface modes due to finite conductivity of the ground plane.

LIST OF PRINCIPAL SYMBOLS

B = Magnetic flux-density, webers/m^2.
d = Thickness of the dielectric layer, m.
E = Electric force, volts/m.
J = Impressed current density, amp/m^2.
l = Quasi-surface current, amp/m.
$M(\theta)$ = Relative magnitude of the θ-component of the electric force with respect to its maximum value.
n = Refractive index, or the square root of the relative permittivity of the dielectric layer with respect to that of the space above the dielectric layer.
P = Power, watts.
T = Impressed polarization density, coulombs/m^2.
r, Φ, z = Cylindrical co-ordinates, m, rad, m.
R, θ, Φ = Spherical co-ordinates, m, rad, rad.
R^e = Radiation resistance, ohms.
$v = (\epsilon\mu)^{-\frac{1}{2}}$ = Characteristic velocity of the half-space above the dielectric layer, m/sec.
z' = z-co-ordinate of the dipole source, m.
$p(z')$ = Vertical Hertzian dipole at $r = 0$, $z = z'$, coulomb-m.
$m(z')$ = Vertical magnetic dipole at $r = 0$, $z = z'$, amp/m^2.
α = Attenuation coefficient, nepers/m.
$\beta = \omega\sqrt{(\epsilon\mu)}$ = Characteristic phase-change coefficient of the half space above the dielectric layer, rad/m.
$n\beta$ = Characteristic phase-change coefficient of the dielectric region, rad/m.
$\delta(z - z')$ = Dirac delta-function.
ϵ = Permittivity of the half space above the dielectric layer, farad/m.
λj = Cylindrical radial (r-directed) propagation co stant of a surface guided mode, rad/m.
μ = Magnetic permeability of the dielectric layer a the space above it, henrys/m.
σ = Conductivity, mhos.
ψ = z-component of the electric Hertz vector, volt-
ψ_m = z-component of the magnetic Hertz vect volt-sec.
$\omega = 2\pi \times$ frequency, rad/sec.
∇^2 = Laplace operator.
$\mathscr{I}$ = Imaginary part of a complex quantity.
$\mathscr{R}$ = Real part of a complex quantity.

Superscript F denotes the far-zone component of the quantity which the superscript is attached.
Superscript G denotes that the quantity is characteristic of surface-guided wave.
Superscript r denotes that the quantity is characteristic of spherical or radiated wave.
Subscript A indicates that the source is in the space above dielectric layer.
Subscript B indicates that the source is in the dielectric layer.
Subscript C indicates that the source is an Abraham dipole ly on the conductor.
Subscript j refers to the jth guided mode.

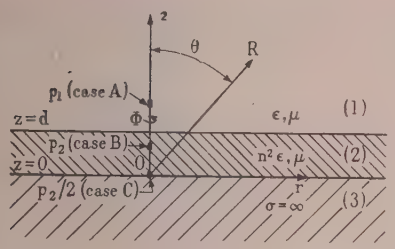
(1) INTRODUCTION

A study of the coated conductor problem was undertaken an effort to explain certain characteristics of an antenna fi pattern obtained from measurements taken over an aluminiu ground screen. The effect of the finite conductivity of aluminium did not yield an adequate explanation of the observ characteristics near the ground screen, and subsequent measu ments showed that the phenomenon could be emphasized by addition of a dielectric layer to the ground plane. Since dielectric-coated metallic surface can support a guided wave any thickness of coating, it was deduced that the thin coating aluminium oxide present on the ground screen was responsi for the behaviour.

The idealized problem involves a uniform layer of perf dielectric covering an infinite, perfectly-conducting plane; Hertzian dipole is in one of three positions: Case A, above dielectric; Case B, in the dielectric; or Case C, an Abrah dipole on the plane, Fig. 1. Unintegrated expressions are deriv for the Hertz potentials in and above the dielectric due to b vertical electric and magnetic dipoles. For the special cases vertical electric dipoles in or near fairly thin dielectric coati the integrated forms of the far-zone fields are obtain Marcuvitz[1] has derived a similar expression for the Hertz poten above the dielectric due to a dipole above but near a thin dielec coating, by a somewhat different method. The mathemati details of this problem are similar to those of the vertical dip over a conducting earth.[2,3]

The resulting fields, comprising a radiated-type wave compensating wave) and one or more guided-type waves, ha

Correspondence on Monographs is invited for consideration with a view to publication.
Dr. Brick is in the Division of Applied Science, Harvard University.



g. 1.—Cross-section of the coated perfect conductor showing the co-ordinate system used.

en combined at a typical distance for a polystyrene coating to eld field patterns. Several examples of these are given illus- ating the distortion of the patterns due to dielectric layers. o illustrate further this distortion, expressions are derived for e total power output of the dipole, the ratio of the powers in e guided and radiated waves, and the radiation resistances of e dipoles. Where possible in the illustrative curves of these pressions, comparison curves for metal surfaces without electric coatings are given. In addition, the expression for the tenuation coefficients of the guided waves due to finite but large nductivities of the conducting plane is derived. No attempt made in the paper to compute the attenuation due to con- ictivity of the dielectric. A close approximation to this tenuation is given by Attwood.[4]

(2) FORMULATION

Let an infinitesimal electric Hertzian dipole, a magnetic dipole, an Abraham dipole be oriented along and parallel to the axis of the cylindrical co-ordinate system r, Φ, z, Fig. 1. his system is independent of Φ. Region 1, $d \leqslant z \leqslant \infty$, with nstants ϵ, μ, $\sigma = 0$, is the space above the dielectric; region 2, ith constants $n^2\epsilon$, μ, $\sigma = 0$, $0 \leqslant z \leqslant d$, is the dielectric region; nd region 3, $z \leqslant 0$ is a perfect conductor with $\sigma = \infty$. The electric constant of medium 2 relative to that of medium 1 n^2, where n is the index of refraction.

As is well known, the electromagnetic field excited by electric magnetic sources oriented in the z-direction of a cylindrical o-ordinate system may be derived from the z-components of e electric or magnetic Hertz vectors. Hence, the problem will e formulated in terms of these components.

The problem is subdivided into five divisions according to the ositions of the dipoles and to whether the electric or magnetic se is treated. The harmonic time dependence, $\varepsilon^{j\omega t}$, is assumed roughout and is suppressed in all expressions.

ertical Electric Dipoles.

Case A. A Hertzian dipole above the dielectric at $d \leqslant z' \leqslant \infty$, $= 0$.

$$\underset{(A)}{\nabla^2}\psi_1 + \underset{(A)}{\beta^2}\psi_1 = -\frac{p_1(z')\delta(z-z')\delta(r)}{2\pi\epsilon r} \quad \text{region 1} \quad . \quad (1)$$

$$\underset{(A)}{\nabla^2}\psi_2 + n^2\underset{(A)}{\beta^2}\psi_2 = 0 \quad \text{region 2} \quad . \quad (2)$$

Case B. A Hertzian dipole in the dielectric at $0 < z' \leqslant d$, $= 0$.

$$\underset{(B)}{\nabla^2}\psi_1 + \underset{(B)}{\beta^2}\psi_1 = 0 \quad \text{region 1} \quad . \quad (3)$$

$$\underset{(B)}{\nabla^2}\psi_2 + n^2\underset{(B)}{\beta^2}\psi_2 = -\frac{p_2(z')\delta(z-z')\delta(r)}{2\pi n^2\epsilon r} \quad \text{region 2} \quad . \quad (4)$$

Case C. An Abraham dipole on the ground plane at $z' = 0$, $r' = 0$. An Abraham dipole is half a Hertzian dipole which with its image forms a Hertzian dipole.

$$\underset{(C)}{\nabla^2}\psi_1 + \underset{(C)}{\beta^2}\psi_1 = 0 \quad \text{region 1} \quad . \quad (5)$$

$$\underset{(C)}{\nabla^2}\psi_1 + n^2\underset{(C)}{\beta^2}\psi_2 = -\frac{p_2(0)\delta(z)\delta(r)}{4\pi n^2\epsilon r} \quad \text{region 2} \quad . \quad (6)$$

The appropriate boundary conditions are

$$(a) \qquad \frac{\partial\psi_1}{\partial z}(r, d, z') = \frac{\partial\psi_2}{\partial z}(r, d, z') \quad . \quad . \quad . \quad . \quad (7a)$$

$$(b) \qquad \epsilon\psi_1(r, d, z') = n^2\epsilon\psi^2(r, d, z') \quad . \quad . \quad . \quad . \quad (7b)$$

$$(c) \qquad \frac{\partial\psi_2}{\partial z}(r, o, z') = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (7c)$$

(*d*) The Sommerfeld condition of radiation,[5,6] where

$\psi_1(r, z, z')$ is the z-component of the electric Hertz vector in region 1.

$\psi_2(r, z, z')$ is the z-component of the electric Hertz vector in region 2.

The E- and B-fields are determined from these by

$$\begin{Bmatrix} E_1 \\ E_2 \end{Bmatrix} = -\hat{r}\frac{\partial^2}{\partial r\partial z}\begin{Bmatrix} \psi_1 \\ \psi_2 \end{Bmatrix} + \hat{z}\left(\frac{\partial^2}{\partial z^2} + \beta^2\begin{Bmatrix} 1 \\ n^2 \end{Bmatrix}\right)\begin{Bmatrix} \psi_1 \\ \psi_2 \end{Bmatrix} \quad . \quad (8)$$

$$\begin{Bmatrix} B_1 \\ B_2 \end{Bmatrix} = -\hat{\Phi}j\frac{\beta^2}{\omega}\frac{\partial}{\partial r}\begin{Bmatrix} \psi_1 \\ n^2\psi_2 \end{Bmatrix} \quad . \quad . \quad . \quad (9)$$

The upper quantities in the brackets are valid in region 1, and the lower quantities in region 2.

$\beta = \omega\sqrt{(\epsilon\mu)}$.

n^2 = Relative dielectric constant of region 2 with respect to region 1.

$\left.\begin{matrix} p_1(z') \\ p_2(z') \end{matrix}\right\}$ is a vertical Hertzian dipole at z' $\begin{cases} d \leqslant z' \leqslant \infty \\ 0 < z' \leqslant d \end{cases}$

$p_2(0)/2$ is an Abraham dipole at $z' = 0$.

$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{\partial^2}{\partial z^2}$ = The Laplacian operator in the cylindrical co-ordinate system with no Φ-dependence.

$\frac{\delta(z-z')\delta(r)}{2\pi r}$ is the Dirac delta-function for a source at $z = z'$, $r = 0$.

Vertical Magnetic Dipoles.

Case A. A magnetic dipole above the dielectric at $d \leqslant z' \leqslant \infty$, $r' = 0$.

$$\underset{(A)}{\nabla^2}\psi_{m1} + \underset{(A)}{\beta^2}\psi_{m1} = -\frac{\mu m_1(z')\delta(z-z')\delta(r)}{2\pi r} \quad \text{region 1} \quad . \quad (11)$$

$$\underset{(A)}{\nabla^2}\psi_{m2} + n^2\underset{(A)}{\beta^2}\psi_{m2} = 0 \quad \text{region 2} \quad . \quad (12)$$

Case B. A magnetic dipole in the dielectric at $0 < z' \leqslant d$, $r' = 0$.

$$\underset{(B)}{\nabla^2}\psi_{m1} + \underset{(B)}{\beta^2}\psi_{m1} = 0 \quad \text{region 1} \quad . \quad (13)$$

$$\underset{(B)}{\nabla^2}\psi_{m2} + n^2\underset{(B)}{\beta^2}\psi_{m2} = -\frac{\mu m_2(z')\delta(z-z')\delta(r)}{2\pi r} \quad \text{region 2} \quad . \quad (14)$$

It should be noted that the problem of a magnetic dipole placed at $z' = 0$ need not be formulated since, according to the theory of images, it yields fields which are identically zero.

The appropriate boundary conditions are

$$\left.\begin{array}{ll} (a) & \psi_{m1}(r, d, z') = \psi_{m2}(r, d, z') \\ (b) & \dfrac{\partial}{\partial z}\psi_{m1}(r, d, z') = \dfrac{\partial}{\partial z}\psi_{m2}(r, d, z) \\ & \text{subject to } \mu_1 = \mu_2 = \mu. \\ (c) & \psi_{m2}(r, o, z') = 0 \\ (d) & \text{The radiation condition,}^{5,6} \end{array}\right\} \quad . \quad . \quad (15)$$

where

$$\left.\begin{matrix}\psi_{m1}(r, z, z') \\ \psi_{m2}(r, z, z')\end{matrix}\right\} \text{ is the } z\text{-component of the magnetic Hertz vector in } \left\{\begin{matrix}\text{region 1} \\ \text{region 2}\end{matrix}\right.$$

The E- and B-fields are determined from these by

$$\begin{Bmatrix} \boldsymbol{E}_{1m} \\ \boldsymbol{E}_{2m} \end{Bmatrix} = j\omega\hat{\Phi}\frac{\partial}{\partial r}\begin{Bmatrix}\psi_{m1} \\ \psi_{m2}\end{Bmatrix} \quad . \quad . \quad . \quad . \quad (16)$$

$$\begin{Bmatrix} \boldsymbol{B}_{m1} \\ \boldsymbol{B}_{m2} \end{Bmatrix} = \hat{z}\left(\frac{\partial^2}{\partial z^2} + \beta^2\begin{Bmatrix}1 \\ n^2\end{Bmatrix}\right)\begin{Bmatrix}\psi_{m1} \\ \psi_{m2}\end{Bmatrix} + \hat{r}\frac{\partial^2}{\partial r \partial z}\begin{Bmatrix}\psi_{m1} \\ \psi_{m2}\end{Bmatrix} \quad . \quad (17)$$

where the subscript m indicates quantities in the magnetic cases and

$$\left.\begin{matrix}m_1(z') \\ m_2(z')\end{matrix}\right\} \text{ is a vertical magnetic dipole at } z' \left\{\begin{matrix} d \leqslant z' \leqslant \infty \\ o < z' \leqslant d \end{matrix}\right.$$

Since the problem is formulated in cylindrical co-ordinates with Φ-symmetry it is convenient to apply the well-known Fourier–Bessel or Hankel transform pair[7] for Φ-symmetry to eqns. (1)–(6) and (11)–(14).

$$\bar{\psi}(\lambda, z, z') = \int_0^\infty r\psi(r, z, z')J_0(\lambda r)dr \quad . \quad . \quad . \quad (18)$$

$$\psi(r, z, z') = \int_0^\infty \lambda\bar{\psi}(\lambda, z, z')J_0(\lambda r)d\lambda \quad . \quad . \quad . \quad (19)$$

After applying eqn. (18) to eqns. (1)–(6) and (11)–(14), performing an integration by parts, and applying the radiation condition* to the left-hand sides of these equations, the following results are obtained for the left-hand sides of these equations:

For eqns. (1), (3), (5), (11), (13)—

$$\left(\frac{\partial^2}{\partial z^2} - l^2\right)\bar{\psi}_1 \text{ or } m_1$$

For eqns. (2), (4), (6), (12) and (14)—

$$\left(\frac{\partial^2}{\partial z^2} - m^2\right)\bar{\psi}_2 \text{ or } m_2$$

where l^2 and m^2 are defined by

$$l^2 = \lambda^2 - \beta^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

$$m^2 = \lambda^2 - n^2\beta^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

After noting that the application of eqn. (18) to the right-hand sides of eqns. (1), (4), (6), (11) and (14) yields the value 1/2 the following transformed equations are obtained:

$$\left(\frac{\partial^2}{\partial z^2} - l^2\right)\left\{\begin{array}{l} \underset{(A)}{\bar{\psi}_1} = -\dfrac{p_1(z')\delta(z - z')}{2\pi\epsilon} \\ \underset{(B)}{\bar{\psi}_1} = 0 \\ \underset{(C)}{\bar{\psi}_1} = 0 \\ \underset{(A)}{\bar{\psi}_{m1}} = -\dfrac{\mu m_1(z')\delta(z - z')}{2\pi} \\ \underset{(B)}{\bar{\psi}_{m1}} = 0 \end{array}\right\} \quad . \quad (2$$

$$\left(\frac{\partial^2}{\partial z^2} - m^2\right)\left\{\begin{array}{l} \underset{(A)}{\bar{\psi}_2} = 0 \\ \underset{(B)}{\bar{\psi}_2} = -\dfrac{p_2(z')\delta(z - z')}{2\pi n^2\epsilon} \\ \underset{(C)}{\bar{\psi}_2} = -\dfrac{p_2(0)\delta(z)}{4\pi n^2\epsilon} \\ \underset{(A)}{\bar{\psi}_{m2}} = 0 \\ \bar{\psi}_{m2} = -\dfrac{\mu m_2(z')\delta(z - z')}{2\pi} \end{array}\right\} \quad . \quad (2$$

These are the transformed forms of eqns. (1), (3), (5), (11) a (13), and (2), (4), (6), (12) and (14) respectively. They satis boundary conditions (7) (*a*)–(*d*) and (15) (*a*)–(*d*), as can be se from eqns. (18) and (19), assuming that the order of integrati and differentiation, with respect to z, may be interchanged.

According to the physical considerations of these problem it is convenient to substitute the following forms in eqns. (2 and (23). If these satisfy the equations and boundary co ditions, they are the unique solutions of the problems except f an arbitrary constant in time.[8] The subscripts indicate t equation to which each trial solution belongs.

$$\left.\begin{array}{l} \dfrac{\epsilon\underset{(A)}{\bar{\psi}_1}(\lambda, z, z')}{p_1(z')} \\ \dfrac{\underset{(A)}{\bar{\psi}_{m1}}(\lambda, z, z')}{\mu m(z')} \end{array}\right\} = \frac{1}{4\pi l}\varepsilon^{-l|z-z'|} + B(\lambda, z')\varepsilon^{-l(z+z'-2d)}\varepsilon^{-2md} + C(\lambda, z')\varepsilon^{-l(z+z'-2d)} \quad . \quad (2$$

where the first term on the right-hand side is the incident wa or the wave radiated by an isolated source, the second term is wave reflected from the metal surface, and the third term is wave reflected from the dielectric surface.

$$\left.\begin{array}{l} \dfrac{\epsilon\underset{(A)}{\bar{\psi}_2}(\lambda, z, z')}{p_1(z')} \\ \dfrac{\bar{\psi}_{m2}(\lambda, z, z')}{\mu m_1(z')} \end{array}\right\} = D(\lambda, z')\varepsilon^{-[l|d-z'|+m(d-z)]} + E(\lambda, z')\varepsilon^{-[l|d-z'|+m(d+z)]} \quad . \quad (2$$

where the first term is a wave transmitted into the dielectr and the second term is a wave reflected from the metal surfa

* The quantity $\left[\frac{\partial\psi}{\partial r}J_0(\lambda r) - \psi r\lambda J_1(\lambda r)\right]_{r=0}^{r=\infty}$ appears after the integration by parts. This disappears identically at $r = 0$ and at $r = \infty$ the radiation condition[5,6] requires its disappearance.

$$\left.\begin{array}{l} \dfrac{n^2\epsilon\bar{\psi}_2(\lambda, z, z')}{\underset{(B)}{p_2(z')}} \\ \dfrac{n^2\epsilon\bar{\psi}_2(\lambda, z, 0)}{\underset{(C)}{p_2(0)}} \\ \dfrac{\bar{\psi}_{m2}(\lambda, z, z')}{\underset{(B)}{\mu m_2(z')}} \end{array}\right\} = \frac{1}{4\pi m}\varepsilon^{-m|z-z'|} + F(\lambda, z')\varepsilon^{-m(2d-z-z')} + G(\lambda, z')\varepsilon^{-m(z+z')}$$

$z' = 0$ for case C . . . . (26)

where the first term is the incident wave, the second term is a wave reflected from the boundary $z = d$, and the third term is a wave reflected from the metal boundary.

$$\left.\begin{array}{l} \dfrac{n^2\epsilon\bar{\psi}_1(\lambda, z, z')}{\underset{(B)}{p_2(z')}} \\ \dfrac{2n^2\epsilon\bar{\psi}_1(\lambda, z, o)}{\underset{(C)}{p_2(0)}} \\ \dfrac{\bar{\psi}_{m1}(\lambda, z, z')}{\underset{(B)}{\mu m_2(z')}} \end{array}\right\} = H(\lambda, z')\varepsilon^{-l(z-d)} \quad . \quad . \quad (27)$$

where the term on the right-hand side represents a wave radiated from the dielectric surface.

It may be observed that the left-hand sides of eqns. (24)–(27) are the Green's functions for eqns. (22)–(23). The proper choice of the coefficients of the first terms on the right-hand sides of eqns. (24) and (26) guarantees that these equations satisfy the conditions for the Green's functions at $z = z'$.

Substituting eqns. (24)–(27) in their proper boundary conditions (7)(*a*)–(*d*) or (15)(*a*)–(*d*), it may be verified that the trial solutions (24)–(27) indeed do satisfy the boundary conditions. Evaluating the constants from $B(\lambda, z')$ to $H(\lambda, z')$ yields the transformed Hertz potentials. Applying the Fourier–Bessel integral, (19), rearranging the resulting integrals, and using eqn. (28),[9] yield the Hertz potentials, (29)–(38), which are the solutions to eqns. (1)–(6) and (11)–(14).

$$\frac{1}{4\pi}\int_0^\infty \frac{J_0(\lambda r)\varepsilon^{-|z|\sqrt{(\lambda^2-k^2)}}}{\sqrt{(\lambda^2-k^2)}}\lambda d\lambda = \frac{\varepsilon^{-jk\sqrt{(r^2+z^2)}}}{4\pi\sqrt{(r^2+z^2)}} \quad . \quad (28)$$

Vertical Electric Dipoles.

Case A.

$$\underset{(A)}{\psi_1}(r, z, z') = \frac{p_1(z')}{4\pi\epsilon}\left[\frac{\varepsilon^{-j\beta R'}}{R'} - \frac{\varepsilon^{-j\beta R''}}{R''} + 2\int_0^\infty \frac{\cosh md\varepsilon^{-l(z+z'-2d)}}{M(\beta, \lambda, n, d)}\lambda J_0(\lambda r)d\lambda\right]$$

region 1 . (29)

$$\underset{(A)}{\psi_2}(r, z, z') = \frac{p_1(z')}{2\pi\epsilon n^2}\int_0^\infty \frac{\cosh mz\varepsilon^{-l(z'-d)}}{M(\beta, \lambda, n, d)}\lambda J_0(\lambda r)d\lambda$$

region 2 . (30)

Case B.

$$\underset{(B)}{\psi_1}(r, z, z') = \underset{(A)}{\psi_2}(r, z', z)p_2\frac{p_2(z')}{p_1(z')} = \frac{p_2(z')}{2\pi\epsilon n^2}\int_0^\infty \frac{\cosh mz'\varepsilon^{-l(z-d)}}{M(\beta, \lambda, n, d)}\lambda J_0(\lambda r)d\lambda$$

region 1 . (31)

$$\underset{(B)}{\psi_2}(r, z, z') = \frac{p_2(z')}{4\pi n^2\epsilon}\left[\frac{\varepsilon^{-jn\beta R'}}{R'} + \frac{\varepsilon^{-j\beta R^V}}{R^V} - 2\int_0^\infty \frac{(l - m/n^2)\cosh mz'\cosh mz\varepsilon^{-md}}{mM(\beta, \lambda, n, d)}\lambda J_0(\lambda r)d\lambda\right]$$

region 2 . (32)

Case C.

$$\underset{(C)}{\psi_1}(r, z, 0) = \frac{p_2(0)}{4\pi\epsilon n^2}\int_0^\infty \frac{\varepsilon^{-l(z-d)}}{M(\beta, \lambda, n, d)}\lambda J_0(\lambda r)d\lambda$$

region 1 . (33)

$$\underset{(C)}{\psi_2}(r, z, 0) = \frac{p_2(0)}{4\pi n^2\epsilon}\left[\frac{\varepsilon^{-jn\beta R}}{R} - \int_0^\infty \frac{(l - m/n^2)\cosh mz\varepsilon^{-md}\lambda J_0(\lambda r)d\lambda}{mM(\beta, \lambda, n, d)}\right]$$

region 2 . (34)

Vertical Magnetic Dipoles.

Case A.

$$\underset{(A)}{\psi_{m1}}(r, z, z') = \frac{\mu m_1(z')}{4\pi}\left[\frac{\varepsilon^{-j\beta R'}}{R'} - \frac{\varepsilon^{-j\beta R''}}{R''} + 2\int_0^\infty \frac{\sinh md}{N(\beta, \lambda, n, d)}\varepsilon^{-l(z+z'-2d)}\lambda J_0(\lambda r)d\lambda\right]$$

region 1 . (35)

$$\underset{(A)}{\psi_{m2}}(r, z, z') = \frac{\mu m_1(z')}{2\pi}\int_0^\infty \frac{\sinh mz}{N(\beta, \lambda, n, d)}\varepsilon^{-l(z'-d)}\lambda J_0(\lambda r)d\lambda$$

region 2 . (36)

Case B.

$$\underset{(B)}{\psi_{m1}}(r, z, z') = \underset{(A)}{\psi_{m2}}(r, z', z)\frac{m_2(z')}{m_1(z')} = \frac{\mu m_2(z')}{2\pi}\int_0^\infty \frac{\sinh mz'}{N(\beta, \lambda, n, d)}\varepsilon^{-l(z-d)}\lambda J_0(\lambda r)d\lambda$$

region 1 . (37)

$$\underset{(B)}{\psi_{m2}}(r, z, z') = \frac{\mu m_2(z')}{4\pi}\left[\frac{\varepsilon^{-j\beta R'}}{R'} - \frac{\varepsilon^{-j\beta R^V}}{R^V} - 2\int_0^\infty \frac{(l - m)\sinh mz'\sinh mz}{mN(\beta, \lambda, n, d)}\varepsilon^{-md}\lambda J_0(\lambda r)d\lambda\right]$$

region 2 . (38)

where $$M(\beta, \lambda, n, d) = l \cosh md + \frac{m}{n^2} \sinh md \quad . \quad . \quad (39)$$

$$N(\beta, \lambda, n, d) = l \sinh md + m \cosh md \quad . \quad . \quad (40)$$

R is the spherical co-ordinate distance of Figs. 2(*a*) and 2(*c*).

$$= \sqrt{(r^2 + z^2)}$$
$$R' = \sqrt{[r^2 + (z - z')^2]} \quad \text{Fig. 2(a)}$$
$$R'' = \sqrt{[r^2 + (z + z' - 2d)^2]} \quad \text{Fig. 2(a)}$$
$$R^{IV} = \sqrt{[r^2 + (z + z')^2]}$$

It is to be observed in eqns. (29)–(38) that the forms of the integrands of the unintegrated terms indicate that these are

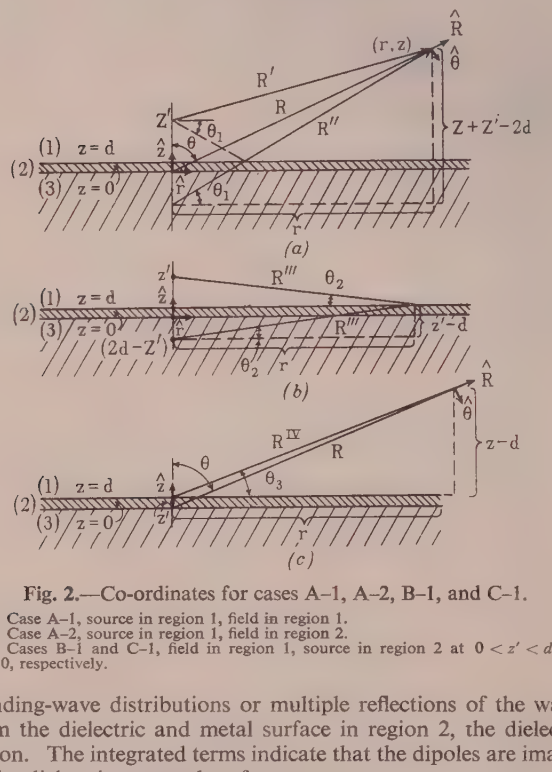


Fig. 2.—Co-ordinates for cases A–1, A–2, B–1, and C–1.
(*a*) Case A–1, source in region 1, field in region 1.
(*b*) Case A–2, source in region 1, field in region 2.
(*c*) Cases B–1 and C–1, field in region 1, source in region 2 at $0 < z' < d$ and $z' = 0$, respectively.

standing-wave distributions or multiple reflections of the waves from the dielectric and metal surface in region 2, the dielectric region. The integrated terms indicate that the dipoles are imaged in the dielectric or metal surface.

The *E*- and *B*-fields may be derived through the application of eqns. (8) and (9) or (16) and (17).

(3) FAR-ZONE FORMULATIONS OF THE HERTZ POTENTIALS

The Hertz potentials ψ_1 and ψ_2 for vertical electric dipoles are asymptotically integrated to obtain far-zone expressions with the restrictions of thin dielectric coatings. The integrations of eqns. (29)–(31) and eqn. (33), Appendix 13.1, are similar to those performed by Tai[10] and others.[11,12] Eqns. (32) and (34) are integrated using the method of Van der Waerden,[13] Appendix 13.2.

The integrals in eqns. (29)–(34) are transformed to a symmetric form for $r > 0$:

$$\int_0^\infty F(\lambda) J_0(\lambda r) \lambda d\lambda = \tfrac{1}{2} \int_{-\infty}^{\infty} F(\lambda) H_0^2(\lambda r) \lambda d\lambda \quad . \quad . \quad (41)$$

through the use of[14]

$$J_0(\lambda r) = \tfrac{1}{2}[H_0^1(\lambda r) + H_0^2(\lambda r)]$$

and $$H_0^2(\lambda r) = -H_0^1(-\lambda r)$$

where $F(\lambda)$ is an even-valued function of λ, and $H_0^2(\lambda r)$ is chose in order to yield outward travelling waves for real positive λ.

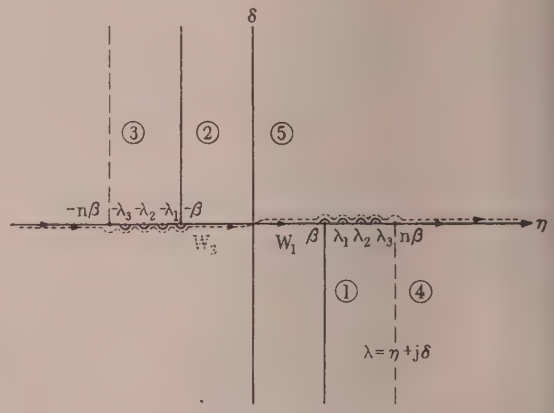


Fig. 3.—Paths of integration in the $\lambda = n + j\delta$ plane.

The integrands in eqns. (29)–(34) have first-order poles whe $M(\beta, \lambda, n, d) = 0$, eqn. (39), or for solutions of

$$\pm n^2 \sqrt{(\lambda^2 - \beta^2)} = \pm \sqrt{(n^2\beta^2 - \lambda^2)} \tan [d\sqrt{(n^2\beta^2 - \lambda^2)}] \quad (4$$

Examining eqn. (42) in the light of eqns. (29)–(31), it is foun that l and m must be either positive real or positive imaginar quantities in order not to violate the radiation condition.[5,6] is also observed that solutions exist only for the same sign on bot sides of eqn. (42) and that the roots, denoted by $\lambda = \lambda_j$, mu satisfy

$$\beta \leqslant \lambda_j \leqslant n\beta \quad . \quad . \quad . \quad . \quad . \quad . \quad (4$$

Fig. 4 contains curves of λ_j/β versus βd, the electrical dielectr thickness, for $n^2 = 2{\cdot}54$, polystyrene, and for $n^2 = 2{\cdot}2$

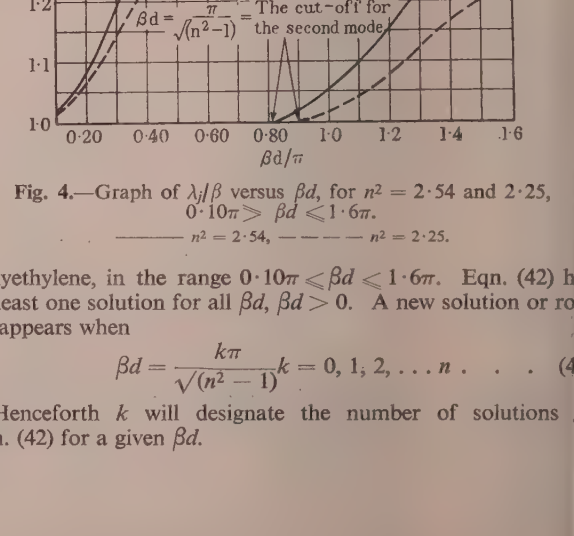


Fig. 4.—Graph of λ_j/β versus βd, for $n^2 = 2{\cdot}54$ and $2{\cdot}25$, $0{\cdot}10\pi \geqslant \beta d \leqslant 1{\cdot}6\pi$.
——— $n^2 = 2{\cdot}54$, — — — — $n^2 = 2{\cdot}25$.

polyethylene, in the range $0{\cdot}10\pi \leqslant \beta d \leqslant 1{\cdot}6\pi$. Eqn. (42) h at least one solution for all βd, $\beta d > 0$. A new solution or ro λ_k appears when

$$\beta d = \frac{k\pi}{\sqrt{(n^2 - 1)}} k = 0, 1, 2, \ldots n \quad . \quad . \quad . \quad (4$$

Henceforth k will designate the number of solutions eqn. (42) for a given βd.

The integrands in eqns. (29)–(31) and (33) have branch points $l = 0$ or $\lambda = \pm\beta$, (1) and (2) in Fig. 3, and at $\lambda = 0$, (5) in g. 3, and poles at $\lambda = \lambda_1, \ldots, \lambda_k$. Upon examination of the tegrals for slightly complex β and n, it is seen that the integrals e to be taken over path W_1 in the complex λ-plane of Fig. 3.

The integrands in eqns. (32) and (34) have additional branch ints at $m = 0$, $\lambda = \pm n\beta$, (3) and (4) in Fig. 3, and in addition e same poles and branch points as the former integrands. ese integrals are to be taken over W_2.

The evaluation of the k residues and the combination of these ith the asymptotic integrations, performed in Appendices 13.1 id 13.2, yield the far-zone Hertz potentials, eqns. (45)–(50), bject to

$$\beta R \gg \left\{\begin{matrix}1\\ \text{and}\\ \beta d\end{matrix}\right\} \quad \text{where} \quad R = \sqrt{(r^2 + z^2)}$$

id

$$\varepsilon^{-R|\beta-\lambda_1|} \ll 0(R^{-2}) \ll 0(R^{-1})$$

$$\varepsilon^{-R|n\beta-\lambda k|} \ll 0(R^{-2}) \ll 0(R^{-1})$$

r in other words the poles are not too near the branch points.

The terms in the summations in eqns. (45)–(50) which are due the evaluation of the residues, are waves which propagate dially and attenuate exponentially in the z-direction. These aves "hug" the surface and are hence called guided or surface aves. They result as eigenfunctions of the configuration, Fig. 1 xcluding $r = 0$).[4,16] The remaining terms will be designated s the radiating or compensating waves.

The far-zone Hertz potentials are:

Case A. Hertz dipole above the dielectric at $d \leqslant z' \leqslant \infty$, $= 0$.

$$\overset{F}{(A)}_1(r, z, z') = \frac{p_1(z')}{\epsilon}\left\{\frac{\varepsilon^{-j\beta R'}}{4\pi R} - \frac{\varepsilon^{-j\beta R''}}{4\pi R''} + \frac{P(\theta_1)}{R''}\varepsilon^{-j\beta R''} + \sum_{J=1}^{k}\frac{A_J(\theta_1)Q(\lambda_J, \beta, n, d)}{\sqrt{(2\pi\lambda_J r)}}\varepsilon^{[-j\lambda_J r-(z+z'-2d)\sqrt{(\lambda_J^2-\beta^2)}]}\right\}$$

$$d \leqslant z \leqslant \infty \qquad \ldots \quad (45)$$

$$\overset{F}{(A)}_2(r, z, z') = \frac{p_1(z')}{n^2\epsilon}\left\{\frac{P(\theta_2)\cos[\beta z\sqrt{(n^2-\cos^2\theta_2)}]}{R'''\cos[\beta d\sqrt{(n^2-\cos^2\theta_2)}]}\varepsilon^{-j\beta R'''} + \sum_{J=1}^{k}\frac{A_J(\theta_2)Q(\lambda_J, \beta, n, d)\cos[z\sqrt{(n^2\beta^2-\lambda_J^2)}]}{\sqrt{(2\pi\lambda_J r)}\cos[\sqrt{d})(n^2\beta^2-\lambda_J^2)]}\varepsilon^{-j\lambda_J r-(z'-d)\sqrt{(\lambda_J^2-\beta^2)}]}\right\} \quad . \quad (46)$$

$$0 \leqslant z \leqslant d$$

Case B. Hertz dipole in the dielectric at $0 \leqslant z \leqslant d$, $r' = 0$.

$$\overset{F}{(B)}_1(r, z, z') = \frac{p_2(z')}{n^2\epsilon}\left\{\frac{P(\theta_3)\cos[\beta z'\sqrt{(n^2-\cos^2\theta_3)}]}{R^{IV}\cos[\beta d\sqrt{(n^2-\cos^2\theta_3)}]}\varepsilon^{-j\beta R^{IV}} + \sum_{J=1}^{k}\frac{A_J(\theta_3)Q(\lambda_J, \beta, n, d)\cos[z'\sqrt{(n^2\beta^2-\lambda_J^2)}]}{\sqrt{(2\pi\lambda_J r)}\cos[d\sqrt{(n^2\beta^2-\lambda_J^2)}]}\varepsilon^{[-j\lambda_J r-(z-d)\sqrt{(\lambda_J^2-\beta^2)}]}\right\} \quad . \quad (47)$$

$$d \leqslant z \leqslant \infty$$

$$\overset{F}{(B)}_2(r, z, z') = \frac{p_2(z')}{\epsilon n^4}\sum_{J=1}^{k}\frac{Q(\lambda_J, \beta, n, d)\cos[z'\sqrt{(n^2\beta^2-\lambda_J^2)}]\cos[z\sqrt{(n^2\beta^2-\lambda_J^2)}]}{\sqrt{(2\pi\lambda_J r)}\cos^2[d\sqrt{(n^2\beta^2-\lambda_J^2)}]}\varepsilon^{-j\lambda_J r} \quad \ldots \quad (48)$$

$$0 \leqslant z \leqslant d$$

Case C. Abraham dipole on the metal surface at $z' = 0$, $r' = 0$.

$$\overset{F}{(C)}_1(r, z, 0) = \frac{p_2(0)}{2n^2\epsilon}\left\{\frac{P(\theta_3)\varepsilon^{-j\beta R^{IV}}}{R^{IV}\cos[\beta d\sqrt{(n^2-\cos^2\theta_3)}]} + \sum_{J=1}^{k}\frac{A_J(\theta_3)Q(\lambda_J, \beta, n, d)}{\sqrt{(2\pi\lambda_J r)}\cos[d\sqrt{(n^2\beta^2-\lambda_J^2)}]}\varepsilon^{[-j\lambda_J r-(z-d)\sqrt{(\lambda_J^2-\beta^2)}]}\right\}$$

$$d \leqslant z \leqslant \infty \qquad \ldots \quad (49)$$

$$\overset{F}{(C)}_2(r, z, 0) = \frac{p_2(0)}{2n^4\epsilon}\sum_{J=1}^{k}\frac{Q(\lambda_J, \beta, n, d)\cos[z\sqrt{(n^2\beta^2-\lambda_J^2)}]}{\sqrt{(2\pi\lambda_J r)}\cos^2[d\sqrt{(n^2\beta^2-\lambda_J^2)}]}\varepsilon^{-j\lambda_J r}$$

$$0 \leqslant z \leqslant d \qquad \ldots \quad (50)$$

where

$$P(\theta)_m = \frac{-j\sin\theta_m}{2\pi\left\{-j\sin\theta_m + n^{-2}\sqrt{(n^2-\cos^2\theta_m)}\tan[\beta d\sqrt{(n^2-\cos^2\theta_m)}]\right\}}$$

$$Q(\lambda_J, \beta, n, d) = \frac{\epsilon^{-j\pi/4}\sqrt{(\lambda_J^2-\beta^2)(n^2\beta^2-\lambda_J^2)}}{\beta^2(n^2-1) + d\sqrt{(\lambda_J^2-\beta^2)}[n^2(\lambda_J^2-\beta^2) + (\beta^2-\lambda_J^2 n^{-2})]}$$

and $A_J(\theta_m)$ is defined by

$$A_J(\theta_m) = 0 \quad \theta_m > \text{arc cos } \beta/\lambda_J$$
$$A_J(\theta_m) = \tfrac{1}{2} \quad \theta_m = \text{arc cos } \beta/\lambda_J$$
$$A_J(\theta_m) = 1 \quad \theta_m < \text{arc cos } \beta/\lambda_J$$

and $$\theta_1 = \text{arc sin}\frac{z + z' - 2d}{R'''} = \text{arc cos}\frac{r}{R'''}$$

$$R'' = \sqrt{[r^2 + (z + z' - 2d)^2]} \quad \text{[see Fig. 2(}a\text{)]}$$

$$\theta_2 = \text{arc sin}\frac{(z' - d)}{R'''} = \text{arc cos}\frac{r}{R'''}$$

$$R''' = \sqrt{[r^2 + (z' - d)^2]} \quad \text{[see Fig. 2(}b\text{)]}$$

$$\theta_3 = \text{arc sin}\frac{(z - d)}{R^{IV}} = \text{arc cos}\frac{r}{R^{IV}}$$

$$R^{IV} = \sqrt{[r^2 + (z - d)^2]} \quad \text{[see Fig. 2(}c\text{)]}$$

If it is assumed that $\beta z' \ll \beta R'''$, the first term in eqn. (46) vanishes and $A_J(\theta_2) \equiv 1$.

The magnitudes of the fields of the guided modes attenuate with height z in two ways owing to (*a*) the exponential or guiding factor $\exp[-z\sqrt{(\lambda_J^2 - \beta^2)}]$, and to (*b*) the coefficient $A_J(\theta_m)$. The second factor is a function of βr, while the first factor does not depend upon the radius. It may be observed that as βr increases the exponential attenuation takes precedence over the coefficient's attenuation. Fig. 5 contains curves of the minimum radii βr, for each mode, at which the field attenuates to 1% of its maximum value owing to exponential attenuation, at a height βz less than that at which the coefficient $A_J(\theta_m)$ equals its critical value. These βr are plotted versus βd, in the range $0{\cdot}10\pi \leqslant \beta d \leqslant 1{\cdot}6\pi$ and for $n^2 = 2{\cdot}54$ and $2{\cdot}25$. For βr greater than these values, $A_J(\theta_m) \equiv 1$ to within 1% accuracy.

The presence of the factor $A_J(\theta_m)$ indicates that as βr increases the guided wave acquires more power, until at $\beta r = \infty$ the total power allotted to it has been attained. Making the assumption that $A_J(\theta_m)$ is identically unity, for βr greater than the values of Fig. 5, is the same as assuming that the guided wave has acquired its total power at this βr. It is valid to within 0·01% accuracy with respect to the Poynting vector.

Since in the paper the far fields are being investigated, it will be assumed henceforth that βr ($= \beta R \sin\theta$) is greater than the

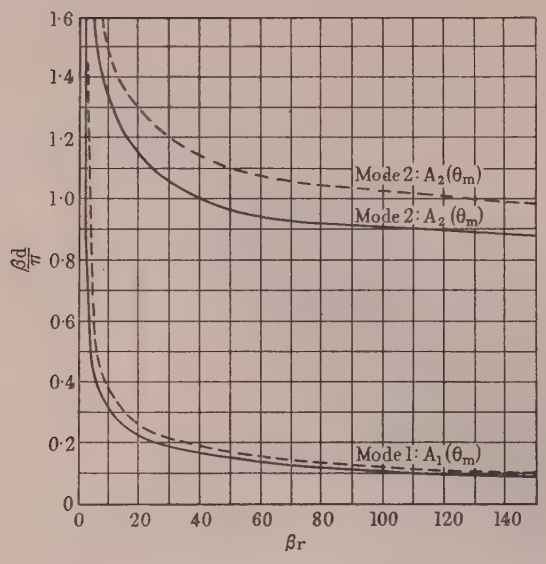


Fig. 5.—Minimum radial distance βr at which the $A_j(\theta_m)$ coefficients of the surface modes are identically unity for all values of z as a function of dielectric thickness, βd, for $0 < \beta d < 1{\cdot}6\pi$.

The curves are based upon an accuracy of 1% of the fields and 0·01% of the Poynting vector.

———— $n^2 = 2{\cdot}54$, — — — — $n^2 = 2{\cdot}25$.

values of Fig. 5 for each βd being investigated. For this reason the notation $A_j(\theta_m)$ will be dropped under the assumption that $A_j(\theta_m) \equiv 1$.

It is to be observed that, in order for the fields to be identical for sources at $z' = d+$ and $z' = d-$, $p_1(d+) = p_2(d-)/n^2$.

(4) THE FAR-ZONE *E*- AND *B*-FIELDS

The representations of the far-zone *E*- and *B*-fields are obtained by applying eqns. (8) and (9) to eqns. (45)–(50). Terms of order higher than r^{-1} are neglected. The assumption is made that $\beta z' \ll \beta R$ so that the radiation term in eqn. (46) may be neglected.

It is convenient to represent the radiating terms in spherical co-ordinates. The following substitutions[15] are made in the radiating waves assuming $\beta R \gg 1$ in Fig. 2.

In the phase factors:

$$\beta R' = \beta R - \beta z' \cos\theta \qquad \beta R'' = \beta R - \beta(2d - z') \cos\theta$$

$$\beta R^{IV} = \beta R - \beta d \cos\theta$$

In the amplitude factors:

$$r/R'' = \cos\theta_1 \simeq \sin\theta = r/R \simeq r/R'$$

$$\frac{z + z' - 2d}{R''} = \sin\theta_1 \simeq \cos\theta = z/R \simeq (z - z')/R'$$

$$\beta R' \simeq \beta R'' \simeq \beta R$$

$$\sin\theta_3 = r/R^{IV} \simeq r/R = \cos\theta$$

$$\cos\theta_3 = r/R^{IV} \simeq r/R = \sin\theta$$

$$\beta R \simeq \beta R^{IV}$$

The substitution of these into eqns. (45)–(50) after eqns. (8) and (9) have been applied leads to the following far-zone fields for vertical dipoles. These are subject to the same restrictions as eqns. (45)–(50) in addition to $\beta z' \ll \beta R$.

Case A. Hertz dipole above the dielectric at $d \leqslant z' \leqslant \infty$, $r' = 0$.

Region 1.

$$\underset{(A)}{\boldsymbol{E}_1^F}(r, z, z') = \underset{(A)}{\boldsymbol{E}_1^{Fr}} + \sum_{J=1}^{k} \underset{(A)}{\boldsymbol{E}_{1J}^{FG}}$$

$$= -V[\hat{R} \times \underset{(A)}{\boldsymbol{B}_1^{Fr}}] + \frac{\omega}{\beta^2} \sum_{J=1}^{k} [-j\sqrt{(\lambda_J^2 - \beta^2)}\, \underset{(A)}{\boldsymbol{B}_{1J}^{FG}} \times \hat{z} - \hat{r} \times \lambda_J \underset{(A)}{\boldsymbol{B}_{1J}^{FG}}] \quad . \quad . \quad . \quad . \quad (51)$$

$$\underset{(A)}{\boldsymbol{B}_1^F}(r, z, z') = \underset{(A)}{\boldsymbol{B}_1^{Fr}} + \sum_{J=1}^{k} \underset{(A)}{\boldsymbol{B}_{1J}^{FG}}$$

$$= -\hat{\Phi} \frac{p_1(z')\beta^3 \sin\theta}{2\pi R\omega\epsilon} \frac{[\cos\theta \cos\zeta + S \sin\zeta]}{(\cos\theta + jS)} \varepsilon^{-j\beta(R - d\cos\theta)}$$

$$- \frac{\hat{\Phi} p_1(z')\beta^2}{\omega\epsilon\sqrt{(2\pi r)}} \sum_{J=1}^{k} \sqrt{(\lambda_J)}\, Q(\lambda_J, \beta, n, d) \varepsilon^{[-j\lambda_J r - (z + z' - 2d)\sqrt{(\lambda_J^2 - \beta^2}} \quad . \quad . \quad . \quad . \quad (52)$$

where $\hat{R}$, $\hat{\theta}$, and $\hat{\Phi}$ are spherical co-ordinates.

$\hat{r}$, $\hat{\Phi}$, and $\hat{z}$ are cylindrical co-ordinates.

$V = (\epsilon\mu)^{-\frac{1}{2}}$ = the characteristic velocity of medium 1.

$\zeta = \zeta(\beta, d, z', \theta) = \beta(d - z') \cos\theta$.

$S = S(n, \theta, d, \beta) = n^{-2}\sqrt{(n^2 - \sin^2\theta)} \tan[\beta d\sqrt{(n^2 - \sin^2\theta)}$

Region 2.

$$\underset{(A)}{\boldsymbol{E}_2^F}(r, z, z') = \sum_{J=1}^{k} \underset{(A)}{\boldsymbol{E}_{2J}^{FG}}$$

$$= \frac{\omega}{n^2\beta^2} \sum_{J=1}^{k} \{-j\sqrt{(n^2\beta^2 - \lambda_J^2)}$$

$$\tan[z\sqrt{(n^2\beta^2 - \lambda_J^2)}] \underset{(A)}{\boldsymbol{B}_{2J}^{FG}} \times \hat{z} - \hat{r} \times \lambda_J \underset{(A)}{\boldsymbol{B}_{2J}^{FG}}\} \quad . \quad (53)$$

$$\underset{(A)}{\boldsymbol{B}_2^F}(r, z, z') = \sum_{J=1}^{k} \underset{(A)}{\boldsymbol{B}_{2J}^{FG}}(r, z, z')$$

$$= -\frac{\hat{\Phi} p_1(z')\beta^2}{\omega\epsilon\sqrt{(2\pi r)}} \sum_{J=1}^{k} \sqrt{(\lambda_J)}\, Q(\lambda_J\beta, n, d) \varepsilon^{[-j\lambda_J r - (z' - d)\sqrt{(\lambda_J^2 - \beta^2}} \frac{\{\cos[z\sqrt{(n^2\beta^2 - \lambda_J^2)}]\}}{\{\cos[d\sqrt{(n^2\beta^2 - \lambda_J^2)}]\}} \quad . \quad (54)$$

Case B. Hertz dipole in the dielectric at $0 < z' \leqslant d$, $r' =$

Region 1.

$$\underset{(B)}{\boldsymbol{E}_1^F}(r, z, z') = \underset{(B)}{\boldsymbol{E}_1^{Fr}}(r, z, z') + \sum_{J=1}^{k} \underset{(B)}{\boldsymbol{E}_{1J}^{FG}}(r, z, z')$$

$$= -V(\hat{R} \times \underset{(B)}{\boldsymbol{B}_1^{Fr}}) + \frac{\omega}{\beta^2} \sum_{J=1}^{k} [-j\sqrt{(\lambda_J^2 - \beta^2)}\, \underset{(B)}{\boldsymbol{B}_{1J}^{FG}} \times z - \hat{r} \times \lambda_J \underset{(B)}{\boldsymbol{B}_{1J}^{Fr}} \quad . \quad . \quad . \quad . \quad (55)$$

$$\underset{(B)}{\boldsymbol{B}_1^F}(r, z, z') = \underset{(B)}{\boldsymbol{B}_1^{Fr}}(r, z, z') + \sum_{J=1}^{k} \underset{(B)}{\boldsymbol{B}_{1J}}(r, z, rz') = -\hat{\Phi} \frac{p_2(z')\beta^3}{2\pi\omega n^2\epsilon R}$$

$$\frac{[\sin\theta \cos\theta]\{\cos[\beta z'\sqrt{(n^2 - \sin^2\theta)}]\varepsilon^{-j\beta(R - d\cos\theta)}\}}{[\cos\theta + jS]\{\cos[\beta d\sqrt{(n^2 - \sin^2\theta)}]\}}$$

$$- \hat{\Phi} \frac{p_2(z')\beta^2}{\omega n^2\epsilon\sqrt{(2\pi r)}} \sum_{J=1}^{k} \sqrt{(\lambda_J)}\, Q(\lambda_J, \beta, n, d) \varepsilon^{[-j\lambda_J r - (z - d)\sqrt{(\lambda_J^2 - \beta^2}}$$

$$\frac{\cos[z'\sqrt{(n^2\beta^2 - \lambda_J^2)}]}{\cos[d\sqrt{(n^2\beta^2 - \lambda_J^2)}]} \quad . \quad (56)$$

Region 2.

$$\underset{(B)}{E_2^{FG}}(r, z, z') = \sum_{J=1}^{k} \underset{(B)}{E_{2J}^{FG}}(r, z, z')$$

$$= \frac{\omega}{n^2\beta^2} \sum_{J=1}^{k} \left\{ -j\sqrt{(n^2\beta^2 - \lambda_J^2)} \tan\left[z\sqrt{(n^2\beta^2 - \lambda_J^2)}\right] \underset{(B)}{B_{2J}^{FG}} \times \hat{z} - \hat{r} \times \lambda_J \underset{(B)}{B_{2J}^{FG}} \right\} \quad . \quad (57)$$

$$\underset{(B)}{B_2^{FG}}(r, z, z') = \sum_{J=1}^{k} \underset{(B)}{B_{2J}^{FG}}(r, z, z')$$

$$= -\hat{\Phi} \frac{p_2(z')\beta^2}{\omega n^2 \epsilon \sqrt{(2\pi r)}} \sum_{J=1}^{k} \sqrt{(\lambda_J)} Q(\lambda_J, \beta, n, d) \varepsilon^{-j\lambda_J r} \frac{\cos\left[z'\sqrt{(n^2\beta^2 - \lambda_J^2)}\right] \cos\left[z\sqrt{(n^2\beta^2 - \lambda_J^2)}\right]}{\cos^2\left[d\sqrt{(n^2\beta^2 - \lambda_J^2)}\right]} \quad . \quad (58)$$

Case C. Abraham dipole on the metal at $z' = 0$, $r' = 0$. The respective quantities are given by the following relations:

$$\underset{(C)}{E^F}(r, z, 0) = \tfrac{1}{2} \lim_{z' \to 0} \underset{(B)}{E^F}(r, z, z') \quad . \quad . \quad . \quad (59)$$

$$\underset{(C)}{B^F}(r, z, 0) = \tfrac{1}{2} \lim_{z' \to 0} \underset{(B)}{B^F}(r, z, z') \quad . \quad . \quad . \quad (60)$$

ith subscripts 1 or 2.

It is to be noted that the radiating or compensating terms e waves of spherical type which attenuate as R^{-1}; the guided rms are waves of cylindrical type which attenuate as $r^{-\frac{1}{2}}$. As or R increases, the ratio of the magnitude of the guided wave o that of the radiated wave increases.

The guided modes are elliptically polarized in region 1 with emi-major axis in the z-direction and semi-minor axis along r. he eccentricity for each mode, which is independent of position,

$$e = \beta/\lambda_J$$

The guided modes are sometimes referred to as "slow waves" ecause their phase velocities, $V_J = \omega/\lambda_J$, are less than the haracteristic velocities of the region 1.

Goubau[16,17] and Attwood[4] have treated, at length, the field nd flux distributions for waves of this type.

(5) FIELD PATTERNS

The field patterns are curves of the relative magnitudes (with espect to their maxima) of the θ-components of the E-fields in egion 1, as functions of θ, in the range of θ between 0 and $\pi/2$, or $\beta R \gg 1$. The dielectric is assumed to be thin so that $\theta \simeq \pi/2$ long the surface between region 1 and 2, Fig. 2. To obtain the ormulae from which these curves are computed, it is necessary o convert the guided components of the fields, eqns. (51)–(60), o spherical co-ordinates and to recombine the fields at some arge fixed βR.

It is important to notice that the field patterns defined above re those measured by a receiving antenna polarized and ravelling in the θ-direction at the chosen value of βR. Since he guided waves have R-components, the patterns are not the elative magnitudes of the total field along the great circle of adius βR. It is also important to observe the manner in which hese patterns vary with βR. Since the radiated wave disappears t $\theta \simeq \pi/2$, while for most finite dielectric thicknesses, βd, the urface wave attenuates rapidly with decreasing θ, the region of nterference between the two is very small. For these reasons, he principal effect of varying βR is to vary the relative magnitudes of the sections of the field patterns, at θ near $\pi/2$ and at θ less than $\pi/2$, by a factor $R^{\frac{1}{2}}$.

Figs. 6–14 contain field patterns for cases A, B, and C, $n^2 = 2{\cdot}54$, $\beta R \simeq 239$ and various values of βd and $\beta z'$, on semi-logarithmic scales, with comparison patterns for dipoles above

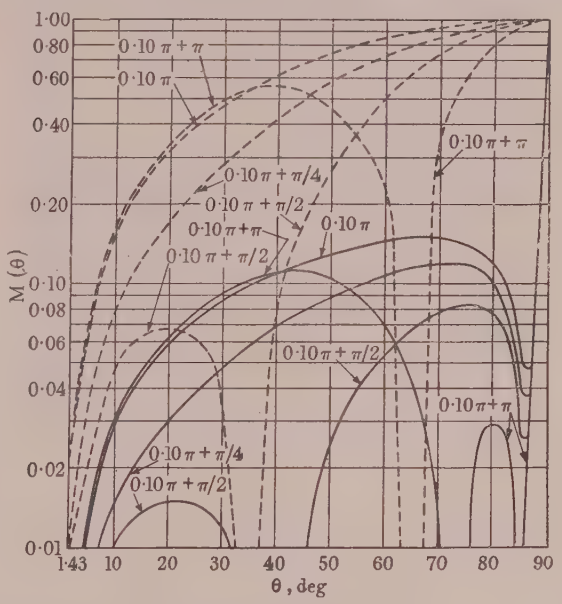


Fig. 6.—Field patterns for case A ($\beta d = 0{\cdot}10\pi$) and comparative with field patterns for $\beta d = 0$.

$\beta R_0 \simeq 239$; $n^2 = 2{\cdot}54$.
The curves, which are marked accordingly, are plotted for $\beta z' = 0{\cdot}10\pi$, $\pi/4 + 0{\cdot}10\pi$; $\pi/2 + 0{\cdot}10\pi$; and $\pi + 0{\cdot}10\pi$.
——— $\beta d = 0{\cdot}10\pi$; — — — — $\beta d = 0$.

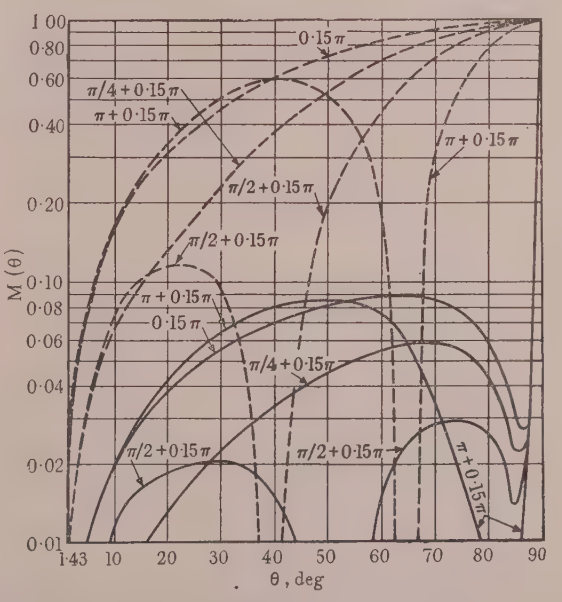


Fig. 7.—Field patterns for case A ($\beta d = 0{\cdot}15\pi$) and comparative patterns for $\beta d = 0$.

$\beta R_0 \simeq 239$; $n^2 = 2{\cdot}54$.
$\beta z' = 0{\cdot}15\pi$; $\pi/4 + 0{\cdot}15\pi$; $\pi/2 + 0{\cdot}15\pi$; $\psi + 0{\cdot}15\pi$.
——— $\beta d = 0{\cdot}15\pi$; — — — — $\beta d = 0$.

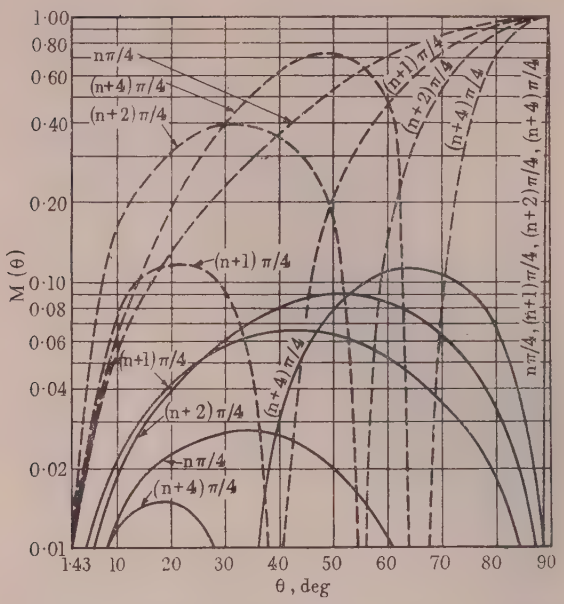


Fig. 8.—Field patterns for case A ($\beta d = n\pi/4$) and comparative patterns for $\beta d = 0$.

$\beta R_0 \simeq 239$; $n^2 = 2{\cdot}54$.
The curves, which are marked accordingly, are plotted for $\beta z' = n\pi/4$; $(n+1)\pi/4$; $(n+2)\pi/4$; and $(n+4)\pi/4$.
——— $\beta d = n\pi/4$; — — — — $\beta d = 0$.

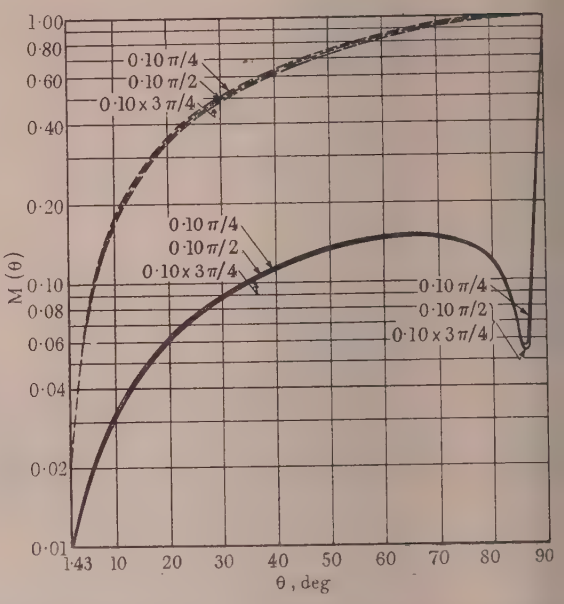


Fig. 10.—Field patterns for case B ($\beta d = 0{\cdot}10\pi$) and comparat patterns for $\beta d = 0$.

$\beta R_0 \simeq 2{\cdot}39$; $n^2 = 2{\cdot}54$.
The curves, which are marked accordingly, are plotted for $\beta z' = 0{\cdot}10$ $0{\cdot}10\pi/2$; and $0{\cdot}10 \times 3\pi/4$.
——— $\beta d = 0{\cdot}10\pi$; — — — — $\beta d = 0$.

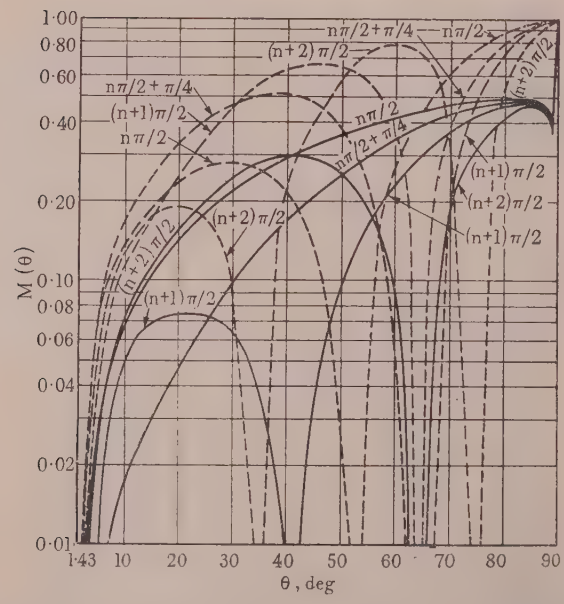


Fig. 9.—Field patterns for case A ($\beta d = n\pi/2$) and comparative patterns for $\beta d = 0$.

$\beta R_0 \simeq 239$; $n^2 = 2{\cdot}54$.
The curves, which are marked accordingly, are plotted for $\beta z' = n\pi/2$; $(n\pi/2) + \pi/4$ $(n+1)\pi/2$; $(n+2)\pi/2$.
——— $\beta d = n\pi/2$; — — — — $\beta d = 0$.

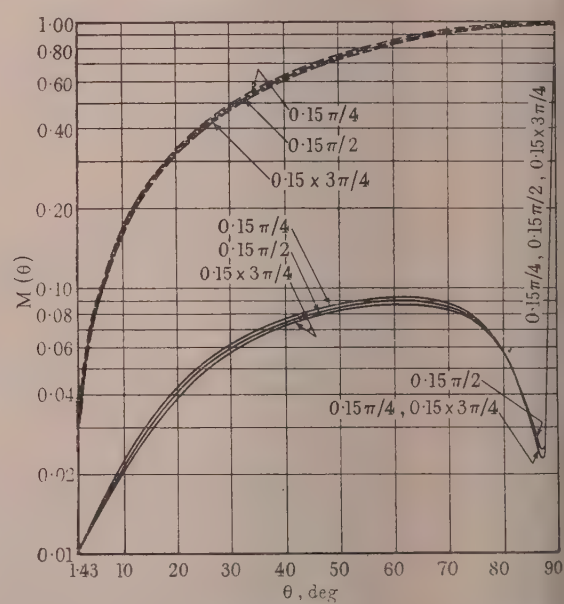


Fig. 11.—Field patterns for case B ($\beta d = 0{\cdot}15\pi$) and comparat patterns for $\beta d = 0$.

$\beta R_0 \simeq 239$; $n^2 = 2{\cdot}54$.
The curves, which are marked accordingly, are plotted for $\beta z' = 0{\cdot}15$ $0{\cdot}15\pi/2$; and $0{\cdot}15 \times 3\pi/4$.
——— $\beta d = 0{\cdot}15\pi$; — — — — $\beta d = 0$.

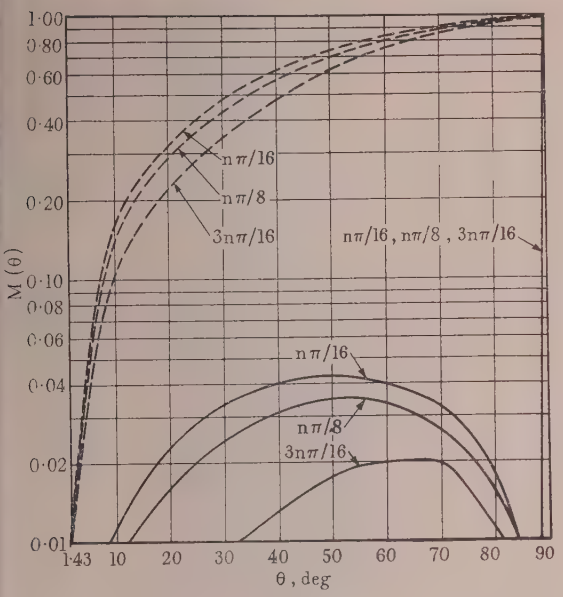


Fig. 12.—Field patterns for case B ($\beta d = n\pi/4$) and comparative patterns for $\beta d = 0$.

$\beta R_0 \simeq 239$; $n^2 = 2{\cdot}54$.
The curves, which are marked accordingly, are plotted for $\beta z' = n\pi/16$; $n\pi/8$; and $3n\pi/16$.
——— $\beta d = n\pi/4$; — — — — $\beta d = 0$.

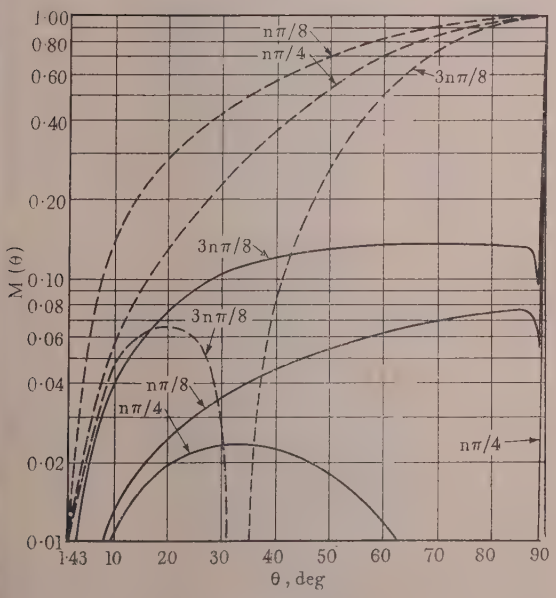


Fig. 13.—Field patterns for case B ($\beta d = n\pi/2$) and comparative patterns for $\beta d = 0$.

$\beta R_0 \simeq 239$; $n^2 = 2{\cdot}54$.
The curves, which are marked accordingly, are plotted for $\beta z' = n\pi/4$; $n\pi/8$; and $n\pi/4$.
——— $\beta d = n\pi/2$; — — — — $\beta d = 0$.

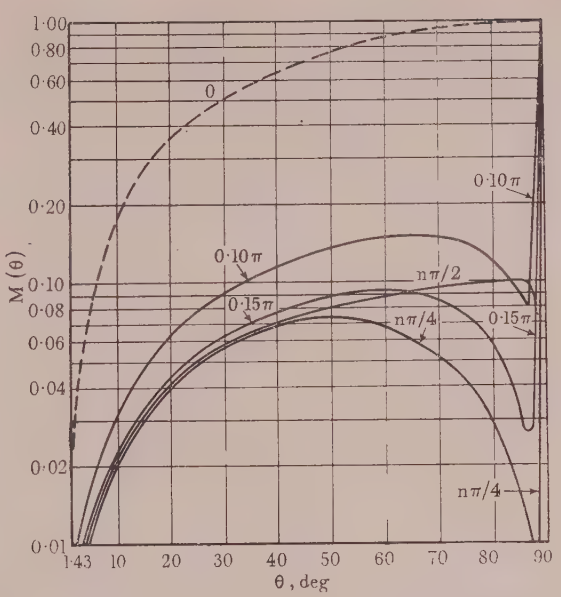


Fig. 14.—Field patterns for case C ($\beta z' = 0$) and comparative patterns for $\beta d = 0$.

$\beta R_0 \simeq 239$; $n^2 = 2{\cdot}54$.
— — — — $\beta d = 0$.
——— $\beta d = 0{\cdot}10\pi$; $0{\cdot}15\pi$; $n\pi/4$; and $n\pi/2$.

or on a perfect uncoated conductor for the same dipole heights $\beta z'$ as those with the coating present.

The dielectric constant of polystyrene was chosen as being typical. $\beta R \simeq 239$ was not only chosen as being typical but was chosen for proposed future experimental use. The thickness $\beta d = n\pi/2$ may seem peculiar, but in Fig. 4 this value is seen to be smaller than, but very close to, $\pi/\sqrt{(n^2 - 1)}$, the value at which multi-mode propagation commences. Since the value $\beta d = n\pi/2$ possesses this quality, besides being convenient for numerical computation, it was chosen. It is clear that the dielectric thicknesses used provide good coverage of the region of single-mode propagation.

The quantities plotted in Figs. 6–14 are

$$M(\theta) = \frac{|E^F_{01}|}{|E^F_{01}|_{\theta = \frac{\pi}{2} = \theta_{max}}}$$

For $\beta d = 0$ and $\beta d \neq 0$.

(6) CONCLUSIONS DRAWN FROM THE FIELD PATTERNS

The main points of interest in Figs. 6–14 are the large maxima or spikes at $\theta = \pi/2$ and the compressions of the rest of the fields with the accompanying minima between these two. These curves are drawn on a logarithmic scale and must be examined closely to appreciate the actual large magnitudes of the maxima at $\theta = \pi/2$ (the surface waves) compared to the rest of the patterns (the radiated waves). The ratios of the powers in the two components of the fields, Figs. 20 and 21, give a better conception of the large size of these maxima, especially if consideration is given to the fact that the surface wave is compressed into a small segment of space, while the radiated wave is dispersed over the complete half-space.

For case *A* the field patterns for $\theta < \pi/2$ are similar to those without the dielectric coating for dipole heights equal to the

dipole heights in case A minus the dielectric thickness. This means that, for case A, the radiated waves are strongly reflected by the dielectric surface.

In cases B and C the effect of the dielectric is to compress the radiated wave into a smaller sector.

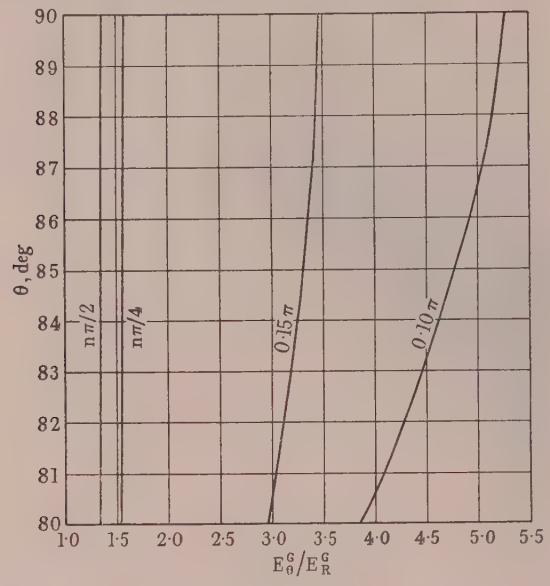


Fig. 15.—Graph of E_θ^G/E_R^G versus θ over the range of significant values of $|E^G|$ for $\beta d = n\pi/2$, $n\pi/4$, $0{\cdot}15\pi$, $0{\cdot}10\pi$ and for $n^2 = 2{\cdot}54$.

Fig. 15 is a plot of $|E_{\theta J}^{FG}|/|E_{RJ}^{FG}|$ versus θ for the βd's of Figs. 6–14 over the range of significant magnitudes of E_J^{FG}. For the cases considered $J = 1$. These yield an indication of the magnitudes of the ratios of θ-components/R-components of the radiation patterns. For the thicker dielectrics these ratios are almost constant, indicating that, since $|E_{zJ}^{FG}|/|E_{rJ}^{FG}|$ is constant and equal to the ratios of θ-components/R-components of Fig. 15 at $\theta = \pi/2$, $E_{\theta J}^{FG}$ and E_{RJ}^{FG} are essentially E_{zJ}^{FG} and E_{rJ}^{FG} respectively over the range of significant values of $|E_J^{FG}|$.

(7) FORMULATION OF THE POWER EQUATION

The purpose of this Section is to formulate the equations which govern the time-average flow of power across a surface surrounding the source $p(z')$. Since the far-zone approximations satisfy the far-zone forms of Maxwell's equations and since the proofs that the powers in the guided- and radiated-type waves are independent require the fields to satisfy the complete forms of Maxwell's equations, it is necessary to formulate the exact near-zone fields.

Consider $\boldsymbol{E}$ and $\boldsymbol{B}$, the exact fields existing in the configuration of Fig. 1, due to a source $p(z')$ in any of three positions A, B, or C of Fig. 1.

For $\beta r > 0$ the transformation of eqns. (29)–(34) to the Hankel function form, eqn. (41), is valid. The exact integrations, if they could be performed, would be taken over paths W_1 and W_2 of Fig. 3. These integrations yield contributions from the poles and branch cuts. The application of eqns. (8) and (9) to the results of the integrations yields the exact E- and B-fields. It is sufficient for the purposes of this Section to represent the branch-cut contributions, $\boldsymbol{E}^r$, $\boldsymbol{B}^r$, symbolically, while the contributions of the poles, $\boldsymbol{E}^G$, $\boldsymbol{B}^G$, may be explicitly evaluated.

For $\beta r > 0$ the exact fields may be represented by

$$(a)\ \boldsymbol{E} = \boldsymbol{E}^r + \boldsymbol{E}^G \qquad (b)\ \boldsymbol{B} = \boldsymbol{B}^r + \boldsymbol{B}^G \qquad (61)$$

$$(a)\ \boldsymbol{E}^G = \sum_{J=1}^{k} \boldsymbol{E}_J^G$$

$$= \sum_{J=1}^{k} \begin{cases} A_{1J}\varepsilon^{-j\pi/2}[\hat{r}\sqrt{(\lambda_J^2 - \beta^2)}H_1^2(\lambda_J r) \\ \qquad + \hat{z}\lambda_J H_0^2(\lambda_J r)]\varepsilon^{-z\sqrt{(\lambda_J^2-\beta^2)}} & \text{region} \\ A_{2J}\varepsilon^{-j\pi/2}\{\hat{r}\sqrt{(n^2\beta^2 - \lambda_J^2)}\sin[z\sqrt{(n^2\beta^2 - \lambda_J^2)}] \\ \qquad H_1^2(\lambda_J r) + \hat{z}\lambda_J \cos[z\sqrt{(n^2\beta^2 - \lambda_J^2)}]H_0^2(\lambda_J r) & \text{region} \end{cases}$$

$$(b)\ \boldsymbol{B}^G = \sum_{J=1}^{k} \boldsymbol{B}_J^G$$

$$= \sum_{J=1}^{k} \hat{\Phi}H_1^2(\lambda_J r)\frac{\beta^2}{\omega} \begin{cases} A_{1J}\varepsilon^{-z\sqrt{(\lambda_J^2-\beta^2)}} & \text{region} \\ A_{2J}n^2 \cos[z\sqrt{(n^2\beta^2 - \lambda_J^2)}] & \text{region} \end{cases}$$

. . . . (6

Where A_J is a function of λ_J, B, n, d, ω, z' and $p(z')$. F $\beta R \gg 1$

$$\left.\begin{aligned} (a)\quad & \lim_{\beta R \gg 1} \begin{Bmatrix} \boldsymbol{E} \\ \boldsymbol{B} \end{Bmatrix} \simeq \begin{Bmatrix} \boldsymbol{E}^F \\ \boldsymbol{B}^F \end{Bmatrix} \\ (b)\quad & \lim_{\beta R \gg 1} \begin{Bmatrix} \boldsymbol{E}^r \\ \boldsymbol{B}^r \end{Bmatrix} \simeq \begin{Bmatrix} \boldsymbol{E}^{Fr} \\ \boldsymbol{B}^{Fr} \end{Bmatrix} \\ (c)\quad & \lim_{\beta R \gg 1} \begin{Bmatrix} \boldsymbol{E}_J^G \\ \boldsymbol{B}_J^G \end{Bmatrix} \simeq \begin{Bmatrix} \boldsymbol{E}_J^{FG} \\ \boldsymbol{B}_J^{FG} \end{Bmatrix} \end{aligned}\right\} \quad . \quad . \quad . \quad (6$$

where the fields on the right-hand sides are given in Section $\boldsymbol{E}^F$ and $\boldsymbol{B}^F$ are approximate solutions to Maxwell's equations f large βr, while $\boldsymbol{E}$, $\boldsymbol{B}$, $\boldsymbol{E}^r$, $\boldsymbol{B}^r$, $\boldsymbol{E}_J^G$, $\boldsymbol{B}_J^G$ are exact solutions Maxwell's equations for $\beta R > 0$.

For $\beta r \gg 1$ the guided or residue waves exist uncancelle but as $\beta r \to 0$ the functional forms of the guided waves becor infinite. Since this is physically impossible except at discre points at $r = 0$, $z = z'$, the so-called radiating or compensati fields, due to the branch cut contributions, must cancel t singular parts of the guided waves as $\beta r \to 0$. An indication the cancellation is given by the appearance of the factor $A_J(\theta$ in the asymptotic results. Despite this cancellation the divisi in eqn. (61) is valid because of the linearity of Maxwell's equatio

Since $p(z')$ is arbitrary, let it be chosen to be a real quanti It may be verified that, for $p(z')$ real, A_{1J} and A_{2J} are real.

The power flowing across any surface S surrounding t source is given by

$$P = \mathscr{R}\left\{\frac{1}{2\mu}\left[\iint_S \hat{n} \cdot (\boldsymbol{E}^r \times \boldsymbol{B}^{r*})dS\right.\right.$$

$$+ \sum_{J=1}^{k} \int_S \hat{n} \cdot (\boldsymbol{E}^r \times \boldsymbol{B}_J^{G*} + \boldsymbol{E}_J^G \times \boldsymbol{B}^{r*})dS$$

$$\left.\left. + \sum_{J=1}^{k}\sum_{m=1}^{k} \int_S \hat{n} \cdot (\boldsymbol{E}_J^G \times \boldsymbol{B}_m^{G*})dS)\right]\right\} \quad . \quad . \quad . \quad . \quad (6$$

where $\hat{n}$ is the outward normal to the volume enclosed by S, a superscript * indicates the complex conjugate.

It will be proved† that

$$\mathscr{R}\left\{\frac{1}{2\mu}\left[\iint_S \hat{n} \cdot (\boldsymbol{E}^r \times \boldsymbol{B}_J^{G*} + \boldsymbol{E}_J^G \times \boldsymbol{B}^{r*})ds\right]\right\} = 0 \quad (6$$

and

$$\mathscr{R}\left[\frac{1}{2\mu}\int_S \hat{n} \cdot (\boldsymbol{E}_J^G \times \boldsymbol{B}_m^{G*})dS = 0 \quad J \neq m\right] \quad . \quad (6$$

† The author is indebted to Dr. George Goubau for suggesting these proofs a recommending references to similar proofs.[18,19]

Some necessary formulae derived by well-known methods[20] are

$$\left[\frac{1}{2\mu}\int_S \hat{n}\cdot(\boldsymbol{E}_A\times\boldsymbol{B}_B^*+\boldsymbol{E}_B\times\boldsymbol{B}_A^*)dS\right]$$
$$=-\mathscr{I}\left[\frac{\omega}{2}\int_V(\boldsymbol{P}_B^*\cdot\boldsymbol{E}_A+\boldsymbol{P}_A^*\cdot\boldsymbol{E}_B)dV\right]$$
$$=-\mathscr{R}\left[\tfrac{1}{2}\int_V(\boldsymbol{J}_B^*\cdot\boldsymbol{E}_A+\boldsymbol{J}_A^*\cdot\boldsymbol{E}_B)dV\right] \quad (67)$$

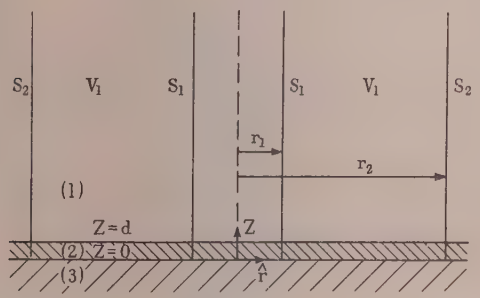
ere S is a closed surface surrounding a volume V.

$\boldsymbol{E}_A$, $\boldsymbol{B}_A$ and $\boldsymbol{E}_B$, $\boldsymbol{B}_B$ are two independent solutions to Maxwell's equation in V.
$\boldsymbol{P}$ and $\boldsymbol{J}$ are the impressed polarization and current densities.
ϵ and μ are assumed to be real in V.
$\hat{n}$ is the outward normal to V.

d the reciprocity condition[21] is

$$\hat{n}\cdot[\boldsymbol{E}_A\times\boldsymbol{B}_B-\boldsymbol{E}_B\times\boldsymbol{B}_A]dS$$
$$=j\omega\mu\int_V(\boldsymbol{P}_A\cdot\boldsymbol{E}_B-\boldsymbol{P}_B\cdot\boldsymbol{E}_A)dV$$
$$=-\mu\int_V\boldsymbol{J}_B\cdot\boldsymbol{E}_A-\boldsymbol{J}_A\cdot\boldsymbol{E}_B)dV \quad . \quad (68)$$

qn. (66) will be proved first.
To this end consider the configuration of Fig. 16 agreeing th Fig. 1.



g. 16.—Concentric cylindrical surfaces of infinite height—S_1 and S_2.

Since the cylinders S_1 and S_2 are semi-infinite and the surface = 0 is a perfect conductor, S may be set equal to S_1 or S_2 eqn. (64). It is therefore sufficient to prove eqn. (66) over or S_2.
If $\boldsymbol{E}_A=\boldsymbol{E}_J^G$, $\boldsymbol{B}_A^*=\boldsymbol{B}_J^{G*}$, $\boldsymbol{E}_B=\boldsymbol{E}_m^G$, and $\boldsymbol{B}_B^*=\boldsymbol{B}_m^{G*}$ are substituted in eqn. (67) with $V=V_1$, Fig. 16, where V_1 is sourceless, e following identity is arrived at:

$$\left[\frac{1}{2\mu}\int_{S_1}\hat{r}\cdot(\boldsymbol{E}_J^G\times\boldsymbol{B}_m^{G*}+\boldsymbol{E}_m^G\times\boldsymbol{B}_J^{G*})dS_1\right]$$
$$=\mathscr{R}\left[\frac{1}{2\mu}\int_{S_2}\hat{r}\cdot(\boldsymbol{E}_J^G\times\boldsymbol{B}_m^{G*}+\boldsymbol{E}_m^G\times\boldsymbol{B}_J^{G*})dS_2\right]$$
$$=\text{constant in } r=K \quad . \quad (69)$$

This is a constant, since r_1 or r_2 may be independently varied hile the other is fixed, or explicitly

$$K=-\mathscr{R}\left\{\frac{1}{\mu}(-j\pi r_{1\text{ or }2}D_{Jm})\right.$$
$$\left.[\lambda_J H_0^2(\lambda_J r_{1\text{ or }2})H_1^1(\lambda_m r_{1\text{ or }2})+\lambda_m H_0^2(\lambda_m r_{1\text{ or }2})H_1^1(\lambda_J r_{1\text{ or }2})]\right\}$$
$$\ldots \quad (70)$$

where the constant D_{Jm}, which is common to both integrals, is given by the following expression

$$D_{Jm}=\frac{\beta^2}{\omega}\left[\int_d^\infty A_{1J}A_{1m}\exp\left\{-z[\sqrt{(\lambda_J^2-\beta^2)}+\sqrt{(\lambda_m^2-\beta^2)}]\right\}dz\right.$$
$$\left.+\int_0^d n^2A_{2J}A_{2m}\cos[z\sqrt{(n^2\beta^2-\lambda_J^2)}]\cos[z\sqrt{(n^2\beta^2-\lambda_m^2)}]dz\right]$$

Specifically, this is true for large βr where the asymptotic forms of the Hankel functions are valid:

$$\lim_{\beta r\gg 1}K=D_{Jm}\left[\sqrt{\left(\frac{\lambda_J}{\lambda_m}\right)}+\sqrt{\left(\frac{\lambda_m}{\lambda_J}\right)}\right]\cos[r(\lambda_J-\lambda_m)] \quad . \quad (71)$$

For $\lambda_J\neq\lambda_m$, this is zero for $r=q\pi/2(\lambda_J-\lambda_m)$; where $q=$ odd integers, but K is a constant, hence the constant must be zero. Therefore D_{Jm} must be zero for all $r>0$. Noting that D_{Jm} is common to each term in eqn. (69), each term of eqn. (69) of the type appearing in eqn. (66) must be separately zero, and hence eqn. (66) is verified.

Eqn. (65) will now be proved.

Consider the configuration of Fig. 17. The source $p(z')$ may

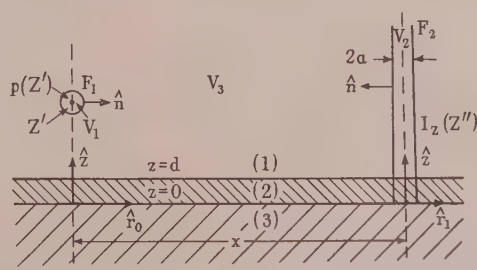


Fig. 17.—Narrow cylinder of infinite height supporting current $I_z(Z'')$, and spherical surface surrounding $p(Z')$ separated by distance x.

be in region 1 or 2 or half-submerged in the conductor at $z=0$, i.e. an Abraham dipole. The current $I(z'')$ on the semi-infinite cylinder V_2 is such as to excite only the Jth guided mode.

After utilizing the representations of the Hankel functions for small arguments, the following representation for $I(z'')$ is arrived at.

$$I_z(z'')=\frac{j4\beta^2}{\omega\lambda_J\mu}\begin{cases}A_{1J}\epsilon^{-z''\sqrt{(\lambda_J^2-\beta^2)}} & d\leqslant z''<\infty\\ A_{2J}n^2\cos[z''\sqrt{(n^2\beta^2-\lambda_J^2)}] & 0\leqslant z''\leqslant d\end{cases}$$
$$\ldots \quad (72)$$

This current excites only $\boldsymbol{E}_J^G(r_1,z)$ and $\boldsymbol{B}_J^G(r_1,z)$. The dipole excites $\boldsymbol{E}(r_0,z)$ and $\boldsymbol{B}(r_0,z)$.

The subscripts for cases A, B, or C agree throughout. It is assumed that $I_z(z'')$ is the total current on F_2 and $p(z')$ is the total polarization in V_1.

The application of eqn. (8) to V_1, F_1, noting that the polarization, $\vec{P}$, is a constant times a delta-function, leads to

$$\int_{F_1} \hat{n} \cdot [\boldsymbol{E} \cdot \boldsymbol{B}_J^G - \boldsymbol{E}_J^G \times \boldsymbol{B}] dF_1 = \omega\mu\lambda_J H_0^2(\lambda_J x)$$

$$\begin{cases} p_1(z')A_{1J}\varepsilon^{-z'\sqrt{(\lambda_J^2-\beta^2)}} & d \leqslant z' \leqslant \infty \\ p_2(z')A_{2J}\cos[z'\sqrt{(n^2\beta^2-\lambda_J^2)}] & 0 < z' < d \\ \dfrac{p_2(0)A_{2J}}{2} & z' = 0 \end{cases} \quad . \quad . \quad . \quad . \quad (73)$$

The application of eqn. (68) to V_2, F_2 in the light of the orthogonality condition, eqn. (66), gives

$$\int_{F_2} \hat{n} \cdot [\boldsymbol{E} \times \boldsymbol{B}_J^G - \boldsymbol{E}_J^G \times \boldsymbol{B}] dF_2 = -\mu \int_0^\infty I(z'') E_z^r(x, z'') dz''$$

$$- \frac{2\beta^2}{\omega} H_0^2(\lambda_J x) \left\{ n^2 A_{2J}^2 d + n^2 \frac{A_{2J}^2 \sin[2d\sqrt{(n^2\beta^2-\lambda_J^2)}]}{2\sqrt{(n^2\beta^2-\lambda_J^2)}} + \frac{A_{1J}^2 \varepsilon^{-2d\sqrt{(\lambda_J^2-\beta^2)}}}{\sqrt{(\lambda_J^2-\beta^2)}} \right\} \quad . \quad (74)$$

The application of eqn. (68) to the sourceless region V_3, noting that the integrations over the surfaces at infinity and at $z = 0$ vanish, and the rearrangement of the resulting equation leads to

$$\mu \frac{\int_0^\infty I(z'') E_z^r dz''}{H_0^2(\lambda_J x)} - \frac{2\beta^2}{\omega} \left\{ \frac{n^2 A_{2J}^2 d + n^2 A_{2J}^2 \sin[2d\sqrt{(n^2\beta^2-\lambda_J^2)}]}{2\sqrt{(n^2\beta^2-\lambda_J^2)}} + \frac{A_{1J}^2 \varepsilon^{-2d\sqrt{(\lambda_J^2-\beta^2)}}}{\sqrt{(\lambda_J^2-\beta^2)}} \right\}$$

$$+ \omega\mu\lambda_J \begin{cases} p_1(z')A_{1J}\varepsilon^{-z'\sqrt{(\lambda_J^2-\beta^2)}} & d \leqslant z' \leqslant \infty \\ p_2(z')A_{2J}\cos[z'\sqrt{(n^2\beta^2-\lambda_J^2)}] & 0 \leqslant z' \leqslant d \\ \dfrac{P_2(0)A_{2J}}{2} & z' = 0 \end{cases} \quad . \quad . \quad . \quad . \quad (75)$$

The right-hand side of eqn. (75) is not a function z or x. Hence, the equation is equal to a constant, G, in x and z.

As βx becomes large, $\beta x \gg 1$; the quantities on the left-hand side of eqn. (75) approach zero in the following manner:

$$\lim_{\beta x \gg 1} E_z^r(x, z'') \sim \frac{1}{\sqrt{(x^2 + z''^2)}}$$

$$\lim_{\beta x \gg 1} H_0^1(\lambda_J x) \sim \frac{1}{\sqrt{x}}$$

Hence $$\lim_{\beta x \to \infty} \frac{G}{\mu} = \lim_{\beta x \to \infty} \frac{\int_0^\infty I(z'') E_z^r dz''}{H_0^2(\lambda_J x)} = 0 \quad . \quad . \quad . \quad (76)$$

Since G is a constant, then $G \equiv 0$ and both sides of eqn. (75) are identically zero.

It is enlightening to observe that the right-hand size of eqn. (75) when set equal to zero offers an alternate method of determining the A_J's. The author has verified the fact that the A_J's are identical when derived by both methods. This offers a check on the paths of integration chosen in Fig. 3.

Since the denominator of the right-hand side of eqn. (76) is well behaved, $\lambda_J x > 0$, the numerator is zero.

With the substitution of the right-hand side of eqn. (72) f $I(z'')$ the numerator on the right-hand side of eqn. (76) become

$$A_{2J} n^2 \int_0^d E_z^r(x, z'') \cos[z''\sqrt{(n^2\beta^2 - \lambda_J^2)}] dz''$$

$$+ A_{1J} \int_d^\infty E_z^r(x, z'') \varepsilon^{-z''\sqrt{(\lambda_J^2-\beta^2)}} dz'' = 0 \quad . \quad (7$$

The following integral is to be considered with reference Fig. 16:

$$\mathscr{R}\left\{ \frac{1}{2\mu} \int_{S_{1 \text{ or } 2}} \hat{n} \cdot [\boldsymbol{E}^r(r, z, z') \times \boldsymbol{B}_J^{G*}(r, z, z')] dS_{1 \text{ or } 2} \right\}$$

$$= -\mathscr{R}\left[\frac{\pi}{\mu} \left[r_{1 \text{ or } 2} \frac{\beta^2}{\omega} H_1^2(\lambda_J r_{1 \text{ or } 2}) \right] \left\{ A_{1J} \int_d^\infty E_z^r \Big|_{r=r_1 \text{ or } r_2} \varepsilon^{-z\sqrt{(\lambda_J^2-\beta^2)}} dz \right.\right.$$

$$\left.\left. + n^2 A_{2J} \int_0^d E_z^r \Big|_{r=r_1 \text{ or } r_2} \cos[z\sqrt{(n^2\beta^2 - \lambda_J^2)}] dz \right\} \right] \quad . \quad (7$$

In the light of eqn. (77), eqn. (78) vanishes.

Consider the sourceless region V of Fig. 16 with the fiel $\boldsymbol{E}$ and $\boldsymbol{B}$ existing in the region. The application of eqn. (67) the two components $\boldsymbol{E}^r$, $\boldsymbol{B}^r$ and $\boldsymbol{E}_J^G$, $\boldsymbol{B}_J^G$ of $\boldsymbol{E}$ and $\boldsymbol{B}$ with t result of eqn. (78) yields

$$\mathscr{R}\left[\frac{1}{2\mu} \int_{S_1} \hat{r} \cdot (\boldsymbol{E}_J^G \times \boldsymbol{B}^{*r}) dS_1 \right] = \mathscr{R}\left[\frac{1}{2\mu} \int_{S_2} \hat{r} \cdot (\boldsymbol{E}_J^G + \boldsymbol{B}^{*r}) dS_2 \right] (7$$

$$= \text{constant in } r$$

since r_1 and r_2 may be independently varied.

The substitution of the explicit expression for E_J^G, allowi βr to become very large and examining the result in a mann similar to that used in deducing that the constant in eqn. (7 was zero, verifies that eqn. (79) is zero for all values of r.

Combining the results of eqns. (78) and (79) verifies eqn. (65

It should be noted that the proof used here shows that ea term in eqn. (65) vanishes. Although a shorter method cou have been used to show that eqn. (65) is zero, it would not ha shown that each term in eqn. (65) vanishes separately or th the A_J's can be determined independently as in eqn. (75).

It follows that eqn. (64) reduces to

$$P = \mathscr{R}\left\{ \frac{1}{2\mu} \left[\int_S \hat{n} \cdot (\boldsymbol{E}^r \times \boldsymbol{B}^{r*}) dS + \sum_{J=1}^{k} \int_s \hat{n} \cdot (\boldsymbol{E}_J^G \times \boldsymbol{B}_J^{G*}) dS \right] \right\}$$

$$= P^r + P^G \quad . \quad (8$$

This allows the division of the power transferred across surface, S, surrounding the source into two or more parts, a hence permits the power radiated and that transferred across by each guided mode to be computed separately.

(8) THE POWER RADIATED BY THE DIPOLES

The time-average power output of the sources may be co puted conveniently by one of two methods:

(*a*) The integration of the Poynting vector over a surface su rounding the source in the far zone.

(*b*) An integration at the source.

Both methods lead to the same results. Method (*b*) is preferr for simplicity.

Let use be made of eqn. (67)[22] with $\boldsymbol{E}_A = \boldsymbol{E}$, $\boldsymbol{B}_B = \boldsymbol{B}$, $\boldsymbol{E}_B =$ $\boldsymbol{B}_A = \boldsymbol{B}$ and $\vec{P}_A = \vec{P}_B = \vec{P}$, noting as in (73) that the polariz

on, $\vec{P}$, has the form of a constant times a delta-function to tain:

$$= \mathscr{R}\left[\frac{1}{2\mu}\int_S \hat{n}\,.\,(\boldsymbol{E}\times\boldsymbol{B})dS\right] = -\mathscr{R}\left[\frac{-j\omega}{2}\int_V \vec{\boldsymbol{P}}^*\,.\,\boldsymbol{E}dV\right]$$

$$= \mathscr{I}\left[\frac{\omega}{2}\begin{cases} p(z')^*\boldsymbol{E}_z(0,z') & \text{Hertzian dipole at } r'=0,\ z'\neq 0 \\ \dfrac{p(0)^*}{2}\boldsymbol{E}_z(0,0) & \text{Abraham dipole } r'=0,\ z'=0 \end{cases}\right] \quad . \quad . \quad . \quad . \quad (81)$$

here S is a surface surrounding the source and V is the volume nclosed by S.

If $\boldsymbol{E}_z(0,z')$ is computed from eqns. (29) to (34) by the application of eqns. (8) and (9), and substituted in eqn. (1), and use is ade of l'Hospital's rule for evaluating indeterminate fractions

ase A.

$$\left[\int_0^\infty I\varepsilon^{-2l(z'-d)}d\lambda\right]$$

$$= \int_0^1 \frac{x^3dx\left\{\sqrt{(1-x^2)}\cos\left[2\beta(z'-d)\sqrt{(1-x^2)}\right] - n^{-2}\sqrt{(n^2-x^2)}\tan\left[\beta d\sqrt{(n^2-x^2)}\right]\sin\left[2\beta(z'-d)\sqrt{(1-x^2)}\right]\right\}}{n^{-4}(n^2-x^2)\tan^2\left[\beta d\sqrt{(n^2-x^2)}\right]+1-x^2}$$

$$+\frac{\pi}{\beta^3}\sum_{J=1}^{k}\varepsilon^{j\pi/4}\lambda_J^2\varepsilon^{-2(z'-d)\sqrt{(\lambda_J^2-\beta^2)}}Q(\lambda_J,\beta,n,d) \quad . \quad (85)$$

Case B.

$$-\mathscr{I}\left[\int_0^\infty \frac{I(lm^{-1}-n^{-2})\varepsilon^{-md}\cosh^2 mz'd\lambda}{n^3\cosh md}\right] = -\frac{\sqrt{(n^2-1)}}{3n} - \frac{\sin\left[2\beta z'\sqrt{(n^2-1)}\right]}{4\beta z'n^3} - \frac{\sqrt{(n^2-1)}}{6n^3}$$

$$+\sqrt{(n^2-1)}\frac{\cos\left[2\beta z'\sqrt{(n^2-1)}\right]}{4n^3\beta^2z'^2} - \frac{\sin\left[2\beta z'\sqrt{(n^2-1)}\right]}{8\beta^3z'^3n^3}$$

$$+\frac{1}{n^3}\int_0^1 \frac{x^3dx\left[\dfrac{\sqrt{(1-x^2)}}{\sqrt{(n^2-x^2)}}-n^{-2}\right]\cos^2\left[\beta z'\sqrt{(n^2-x^2)}\right]\left\{n^{-2}\sqrt{(n^2-x^2)}\tan^2\left[\beta d\sqrt{(n^2-x^2)}\right]+\sqrt{(1-x^2)}\right\}}{n^{-4}(n^2-x^2)\tan^2\left[\beta d\sqrt{(n^2-x^2)}\right]+1-x^2}$$

$$+\frac{\pi}{n^5\beta^3}\sum_{J=1}^{k}\frac{\lambda_J^2\varepsilon^{j\pi/4}\cos^2\left[z'\sqrt{(n^2\beta^2-\lambda_J^2)}\right]Q(\lambda_J,\beta,n,d)}{\cos^2\left[d\sqrt{(n^2\beta^2-\lambda_J^2)}\right]} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (86)$$

Case C.

$$-\mathscr{I}\left[\int_0^\infty \frac{I(lm^{-1}-n^{-2})\varepsilon^{-md}d\lambda}{2n^3\cosh md}\right]$$

$$= -\frac{(2n^2+1)}{6n^3}\sqrt{(n^2-1)} + \frac{1}{2n^3}\int_0^1 \frac{x^3dx\left[\dfrac{\sqrt{(1-x^2)}}{\sqrt{(n^2-x^2)}}-n^{-2}\right]\left\{n^{-2}\sqrt{(n^2-x^2)}\tan^2\left[\beta d\sqrt{(n^2-x^2)}\right]+\sqrt{(1-x^2)}\right\}}{n^{-4}(n^2\beta^2-x^2)\tan^2\left[\beta d\sqrt{(n^2-x^2)}\right]+1-x^2}$$

$$+\frac{\pi}{2n^5\beta^3}\sum_{J=1}^{k}\frac{\lambda_J^2\varepsilon^{j\pi/4}Q(\lambda_J,\beta,n,d)}{\cos^2\left[d\sqrt{(n^2\beta^2-\lambda_J^2)}\right]} \quad . \quad . \quad (87)$$

r the well-known expression for the power radiated by an solated Hertz dipole[23,24] is utilized, the expressions for P result:

Case A.

$$P_{(A)} = \frac{|p_1(z')|^2\omega\beta^3}{4\pi\epsilon}\left\{\frac{1}{3}+\frac{\cos\left[2\beta(z'-d)\right]}{4\beta^2(z'-d)^2} - \frac{\sin\left[2\beta(z'-d)\right]}{8\beta^3(z'-d)^3} + \mathscr{I}\left[\int_0^\infty I\varepsilon^{-2l(z'-d)}d\lambda\right]\right\} \quad . \quad (82)$$

Case B.

$$P_{(B)} = \frac{|p_2(z')|^2\omega n\beta^3}{4\pi\epsilon}\left\{\frac{1}{3}+\frac{\sin 2n\beta z'}{8\beta^3n^3z'^3} - \frac{\cos 2n\beta z'}{4\beta^2n^2z'^2} - \mathscr{I}\left[\int_0^\infty \frac{I(lm^{-1}-n^{-2})\varepsilon^{-md}\cosh^2 mz'}{n^3\cosh md}d\lambda\right]\right\} \quad . \quad (83)$$

Case C.

$$P_{(C)} = \frac{|p_2(0)|^2\omega n\beta^3}{8\pi\epsilon}\left\{\frac{1}{3} - \mathscr{I}\left[\int_0^\infty \frac{I(lm^{-1}-n^{-2})\varepsilon^{-md}}{2n^3\cosh md}d\lambda\right]\right\} \quad . \quad (84)$$

where $$I = \frac{\lambda^3\beta^{-3}}{l+mn^{-2}\tanh md}$$

If the integrals are closely examined and integrated wherever possible, they are seen to result in

where $Q(\lambda_J,\beta,n,d)$ is defined in Section 3.

If method (*a*), in conjunction with eqns. (80) and (51)–(60), were used to compute P^G, it would be found that this yields exactly the last term in eqns. (85)–(87). Hence the terms in the summations when substituted in eqns. (82)–(84) are the total powers transferred into the guided modes from the dipoles. It is to be noted that it is necessary to prove that the power can be subdivided as in eqn. (80) before the two divisions can be compared.

As was previously observed, the total power allotted to a guided mode [the summation terms in eqns. (85)–(87)] is not

acquired by the mode until βr becomes infinite. To a close approximation this is true for βr greater than the values of Fig. 5. To this approximation, then, eqns. (82)–(87) may be given the following interpretation according to eqn. (80); P is the total power transferred across S:

$$P^G = \sum_{J=1}^{k} P_J^G$$

P^G is equal to the summation terms in eqns. (82)–(87) and is equal to the power transferred across S by the guided modes, where S has a minimum radius βR which is greater than the values of βr in Fig. 5. P^r is equal to the remaining terms of eqns. (82)–(87) and is equal to the power transferred across S by the compensating waves:

$$P = P^r + \sum_{J=1}^{k} P_J^G$$

If eqns. (82) and (83) are evaluated at $z' = d+$ and $z' = d-$, respectively, it turns out that the powers are identical if $p_2(d-) = n^2 p_1(d+)$. This verifies the previous result.

Figs. 18(a) and 18(b) contain curves of $4\pi\epsilon P/|p^2|\omega\beta^3$ as a function of dipole height, $\beta z'$, for $n^2 = 2{\cdot}54$, for two thicknesses of

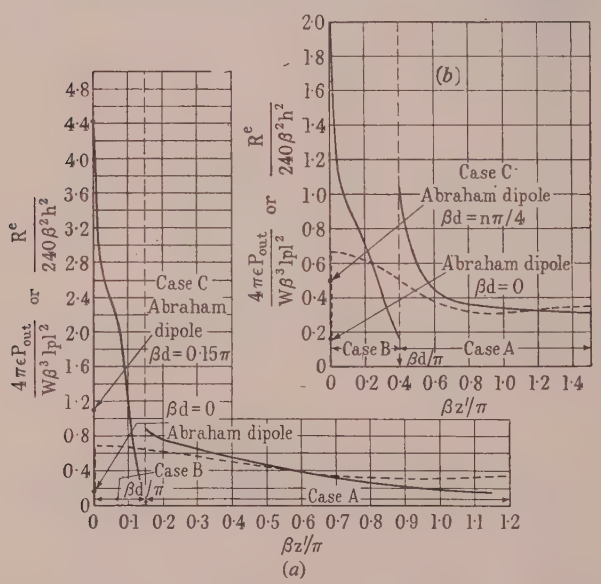


Fig. 18.—Total power radiated and radiation resistance as a function of $\beta z'$.

(a) $n^2 = 2{\cdot}54$.
——— $\beta d = 0{\cdot}15\pi$; - - - - - - $\beta d = 0$.
(b) $n^2 = 2{\cdot}54$.
——— $\beta d = n\pi/4$; - - - - - - $\beta d = 0$.

dielectric, $\beta d = 0{\cdot}15\pi$ and $n\pi/4$. The integrals in eqns. (85)–(87) were computed by numerical integration. Comparison curves with $\beta d = 0$ are given.

The discontinuities at $\beta z = \beta d$ in Figs. 18(a) and 18(b) are due to the above-mentioned discontinuities of the magnitudes of the sources. The total powers are seen to decrease with $\beta z'$ in the dielectric. The slight hump in this region in Fig. 18(b) is due in part to the maximum of P^G as seen in Fig. 20. For $\beta z' > \beta d$ the powers fluctuate around the values for $\beta z' = 0$. Both converge upon 1/3 as $\beta z'$ approaches large values, indicating that the coupling to the guided modes and image fields approaches zero for large $\beta z'$.

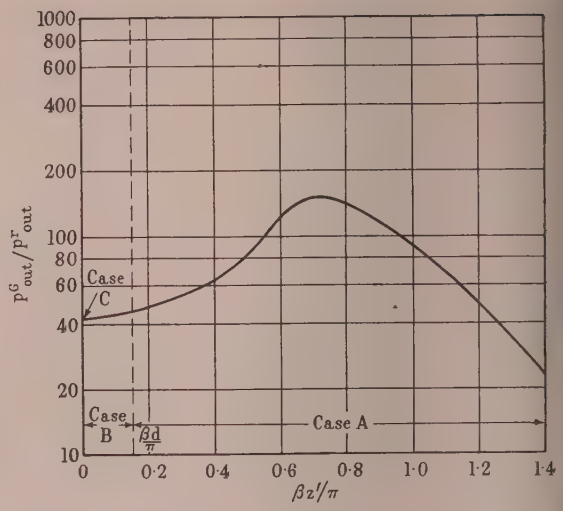


Fig. 19.—P^G_{out}/P^r_{out} as a function of $\beta z'$ for $\beta d = 0{\cdot}15\pi$.
$n^2 = 2{\cdot}54$.
This includes cases A, B and C as $\beta z'$ varies.

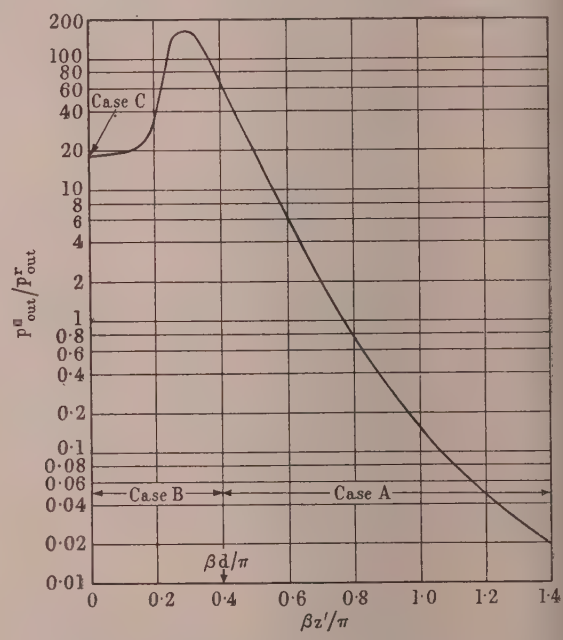


Fig. 20.—P^G_{out}/P^r_{out} as a function of $\beta z'$ for $\beta_d = n\pi/4$.
$n^2 = 2{\cdot}54$.
This includes cases A, B and C as $\beta z'$ varies.

Figs. 19 and 20 contain curves of P^G/P^r versus dipole heigh $\beta z'$, for the same βd's and n as Figs. 18(a) and 18(b). Figs. 19 an 20 are seen to have maxima near the dielectric surface and t decrease uniformly with $\beta z'$ beyond these maxima. In Fig. 2 the maximum is in the dielectric region, while as βd decreases moves into the region above the dielectric as in Fig. 19.

For the smaller $\beta z'$, the large magnitudes of the ratios i Figs. 19 and 20 indicate that an overwhelming portion of th power radiated by the dipole is transferred to the guided c surface waves. In these cases the nomenclature of "compen sating wave" for the radiated component of the field is a preferre one.

THE ATTENUATION COEFFICIENT OF A GUIDED MODE DUE TO THE FINITE CONDUCTIVITY OF THE GROUND PLANE

The attenuation coefficient, α_J, for $\beta r \gg 1$, of the Jth guided mode due to a large but finite conductivity of the ground plane may be derived by the use of the following formula:[25]

$$\alpha_J = -\frac{1}{2}\frac{\partial P_J^G/\partial_r}{P_J^G} \quad . \quad . \quad . \quad . \quad . \quad (88)$$

If the following procedure is followed, eqn. (89) will result for cases A, B, and C.

(*a*) Assume the fields are those of eqns. (51)–(60) except for a small component $Er_J^{FG}(r, 0)$.

(*b*) Assume P_J^G is that of eqns. (82)–(87).

(*c*) Set $H_{\Phi J}^G(r, 0) \simeq -l_{rJ}(r, 0) \simeq -Er(r, 0)/Z_S$, where $l_{rJ}(r, 0)$ is a quasi-surface current.

$$Z_S = (1 + j)\sqrt{\left(\frac{\omega\mu}{2\sigma}\right)}$$

(*d*) Compute

$$-\frac{\partial P_J^G}{\partial r} = \frac{1}{2}\mathscr{R}\left[\int_0^{2\pi} Z_S |l_{rJ}(r, 0)|^2 r d\Phi\right]$$

$$= \frac{R_S}{2}\int_0^{2\pi} |H_{\Phi J}(r, 0)|^2 r d\Phi$$

(*e*) Substitute d and b in eqn. (88)

$$= \frac{n^2\beta^2}{\lambda_J\sqrt{(2\omega\sigma\mu)}} \frac{\sqrt{(\lambda_J^2 - \beta^2)}[(\beta^2 - n^{-2}\lambda_J^2) + n^2(\lambda_J^2 - \beta^2)]}{\{\beta^2(n^2 - 1) + d\sqrt{(\lambda_J^2 - \beta^2)}[(\beta^2 - n^{-2}\lambda_J^2) + n^2\sqrt{(\lambda_J^2 - \beta^2)}]\}} \quad . \quad . \quad . \quad . \quad (89)$$

for cases A, B, and C.

Fig. 21 contains a plot of $\alpha_J\sqrt{(2\omega\sigma\mu)}/n^2\beta^2$ as βd varies, for $n^2 = 2{\cdot}54$, and $0{\cdot}10\pi \leqslant \beta d \leqslant 1{\cdot}6\pi$, for the existing modes 1 and 2. It is to be noted that the attenuation coefficient for each mode increases to a maximum and then decreases uniformly with increasing βd. These maximum values of attenuation appear approximately halfway between the cut-off thicknesses. The attenuation coefficients for the higher modes are lower than those for the lower modes since the factor d appears in the denominator of (89) and the higher modes appear with larger βd.

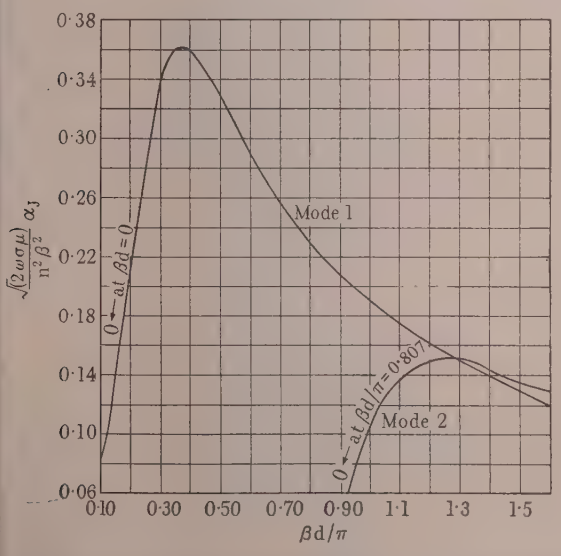


Fig. 21.—Attenuation of a surface mode due to a large but finite conductivity σ of the ground plane as a function of dielectric thickness βd.

$n^2 = 2{\cdot}54$; $0{\cdot}10\pi \leqslant \beta d \leqslant 1{\cdot}6\pi$.

(10) RADIATION RESISTANCE

The radiation resistance may be computed by referring it to the uniform current on the dipole, $I(z')$:

$$R^e = \frac{2P}{|I(z')|^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (90)$$

where $I(z') = j\omega p(z')/2h$ for both the Abraham and Hertz dipoles and h = the half-length of the Hertz dipole = the length of the Abraham dipole. Hence

$$\frac{R^e}{240\beta^2h^2} = \frac{4\pi\epsilon P}{\omega\beta^3|p(z')|^2} \quad . \quad . \quad . \quad . \quad (91)$$

This is plotted with P in Figs. 18 and 19.

(11) ACKNOWLEDGMENTS

The author wishes to thank Prof. R. W. P. King and Dr. Tetsu Morita for their advice and encouragement. He also wishes to express his gratitude to Mrs. Mary Andrews for carrying out the extensive computations.

The work described was made possible through support extended to Cruft Laboratory, Harvard University, jointly by the Navy Department (Office of Naval Research), the Signal Corps of the U.S. Army, and the U.S. Air Force.

(12) REFERENCES

(1) MARCUVITZ, N.: "Guided Wave Concept in Electromagnetic Theory," *Research Report R–269–52* (P.I.B.–208, ONR, May 14, 1952. Polytechnic Institute of Brooklyn, Microwave Research Institute), p. 1.

(2) SOMMERFELD, A.: "Partial Differential Equations in Physics" (Academic Press, New York, 1949), p. 246.

(3) STRATTON, J. A.: "Electromagnetic Theory" (McGraw-Hill, New York, 1941), p. 573.

(4) ATTWOOD, S.: "Surface Wave Propagation over a Coated Plane Conductor," *Journal of Applied Physics*, 1951, **22**, p. 504.

(5) FRANK-MISES: "Differentialgleichungen der Physik" (Friedrich Vieweg und Sohn, Brunswick, Germany, 1935), Vol. II, Chap. XIX, Sec. 5.

(6) KING, R. W. P.: "Electromagnetic Engineering" (McGraw-Hill, New York, 1945), Vol. II, p. 232.

(7) STRATTON, J. A.: *loc. cit.*, p. 371.

(8) STRATTON, J. A.: *ibid.*, pp. 486–487.

(9) STRATTON, J. A.: *ibid.*, pp. 575–576.

(10) TAI, C. T.: "The Effect of a Grounded Slab on Radiation from a Line Source," *Journal of Applied Physics*, 1951, **22**, p. 405.

(11) STRATTON, J. A.: *loc. cit.*, p. 584.

(12) KARP and SOLLFREY: "Diffraction by a Dielectric Wedge," *Research Report* No. EM–23 (Mathematical Research Group, New York University, October, 1950), pp. 19 and 31.

(13) VAN DER WAERDEN, B. L.: "On the Method of Saddle Points," *Applied Scientific Research*, **B–2**, No 7, p. 43.

(14) JAHNKE and EMDE: "Tables of Functions" (Dover, New York), p. 134.
(15) KING, R. W. P.: *loc. cit.*, pp 268 and 276.
(16) GOUBAU, G : "Surface Waves and Their Application to Transmission Lines," *Journal of Applied Physics*, 1950, **21**, p. 1119.
(17) GOUBAU, G:. "Single Conductor Surface-Wave Transmission Lines," *Proceedings of the Institute of Radio Engineers*, 1951, **39**, pp. 619–624.
(18) GOUBAU, G.: "On the Excitation of Surface Waves," *ibid.*, 1952, **40**, p. 865.
(19) GOUBAU, G.: "Über die Zennecksche Bodenwelle," *Zeitschrift für angewande Physik*, 1951, **3**, p. 103.
(20) KING, R. W. P.: *loc. cit.*, p. 220.
(21) STRATTON, J. A.: *loc. cit.*, p. 479.
(22) KING, R. W. P.: *loc. cit.*, p. 221.
(23) STRATTON, J. A.: *loc. cit.*, p. 435.
(24) KING, R. W. P.: *loc. cit.*, p. 296.
(25) LAMONT, H. R. L.: "Wave Guides" (John Wiley, New York, 1949), p. 45.

(13) APPENDICES

(13.1) Illustration of the Type of Integration performed to yield Eqns. (45)–(47) and (49)

To illustrate the type of integration performed to yield eqns. (45)–(47) and (49), consider the integral in eqn. (29), after it has been transformed by eqn. (41), to be integrated over W_1 of Fig. 3.

$$I = \frac{1}{4\pi}\int_0^\infty \frac{\varepsilon^{-l(z+z'-2d)}}{l + mn^{-2}\tanh md} H_0^2(\lambda r)\lambda d\lambda \quad . \quad . \quad (92)$$

It is convenient to make the following substitution:

$$R'' \sin\theta_1 = z + z' - 2d;\ R'' \cos\theta_1 = r \quad \text{[see Fig 2(a)]} \quad . \quad (93)$$

$$\lambda = \beta \sin\alpha;\ \sqrt{(\beta^2 - \lambda^2)} = \beta\cos\alpha \quad . \quad . \quad (94)$$

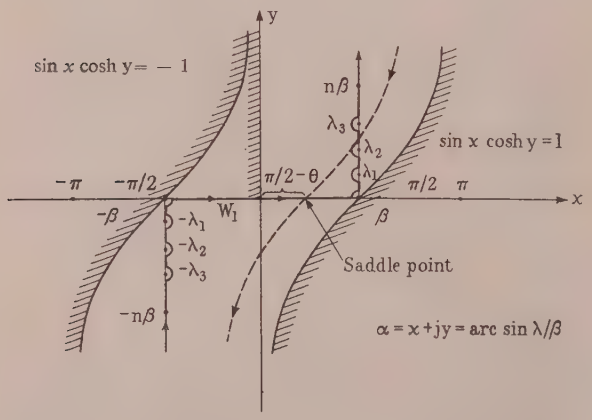


Fig. 22.—Path of integration in the α = arc sin λ/β-plane.
— — — — Path of asymptotic integration w, defined by sin $(x + \theta)$ cosh $y = 1$.

The Riemann sheet of interest in the λ-plane of Fig. 3 transforms into the curved strip in the α-plane, Fig 22. The transformation of the branch cuts (which are not branch cuts in the α-plane) are defined by sin x cosh $y = \pm 1$. Making the assumption that $\beta R'' \cos\theta_1 \gg 1$, so that H_0^2 may be replaced by its asymptotic form for large argument, and substituting eqns (9 and (94) in eqn. (92), (92) becomes

$$I = -\frac{\varepsilon^{j\pi/4}\beta^2}{4\pi}\int_{W_1}\sqrt{\left(\frac{2\sin\alpha}{\pi R''\beta\cos\theta_1}\right)} \frac{\cos\alpha\,\varepsilon^{-j\beta R''(\cos\alpha\sin\theta_1 + \sin\alpha\cos\theta_1)}d\alpha}{\{[-j\beta\cos\alpha + \beta n^{-2}\sqrt{(n^2 - \sin^2\alpha)}\tan[\beta d\sqrt{(n^2 - \sin^2\alpha)}} \quad . \quad . \quad . \quad (9$$

This is of the proper form for an integration by the method steepest descents,[10,12] when $R'' \gg d$

The saddle-point is at $\alpha = \pi/2 - \theta$

The path of steepest descent, W, is defined by

$$\sin(x + \theta_1)\cosh y = 1 \text{ (Fig. 2)}$$

If the assumption is made that the poles are not near enou to the value $\alpha = \pi/2$ to affect the integrand before the exponenti factor has attenuated it to negligible value,*[12] the integratic over W results in:

$$+\frac{j\beta\varepsilon^{-j\beta R''}}{2\pi R''} \frac{\sin\theta_1}{-j\beta\sin\theta_1 + \beta n^{-2}\sqrt{(n^2 - \cos^2\theta_1)}\tan[\beta d\sqrt{(n^2 - \cos^2\theta_1}} \quad . \quad . \quad . \quad . \quad (9$$

and it may be observed that

$$I = \int_{W_1} \ldots d\alpha = -\int_W \ldots d\alpha - 2\pi j \sum \text{Residues enclosed by } W \text{ and } W_1 \quad . \quad (9$$

When W passes through a pole, one-half the residue is i cluded[10,12] as a contribution to the integration. If the pole included between W and W_1, the whole residue is contribut to eqn. (97), while if the pole is excluded by path W–W_1, the is no contribution from it.

It is relatively easy to verify that W passes through a po when θ_1 = arc cos β/λ_J, and from this relationship the coefficien $A_J(\theta_1)$ of the residue waves arise. The residue terms are evaluat by well-known methods.

The rapid decrease of the coefficient $A_J(\theta_1)$ at its critical val is due to the fact that for large βR the path passes very rapid through the pole as θ_1 varies.

The results obtained thus far have been restricted $\beta R''\cos\theta_1 \gg 1$ by virtue of the assumption that the Hank function could be replaced by its asymptotic form. Actuall the restriction is less strict for a good approximation and it m be modified to $\beta R'' \gg 1$ and $(\beta R''\cos\theta_1)^2 \gg 1$. This leav only a small region near $\theta_1 = \pi/2$ where the results develop are not valid. The restrictions may be further modified by t following.

Consider eqn. (29) for $r = 0$, $\theta_1 \simeq \pi/2$, $\beta R'' \simeq \beta z \gg 1$.

$$I\Big|_{r=0} = \frac{1}{2\pi}\int_0^\infty \frac{\varepsilon^{-l(z+z'-2d)}}{l + mn^{-2}\tanh md}\lambda d\lambda \quad . \quad . \quad (9$$

Performing the same transformations, eqns. (93) and (94 and integrating over half the path W (the part on and above t x-axis of Fig 22) in the α-plane, the following is obtained:

$$I\Big|_{r=0} = \frac{-j}{2\pi}\frac{\varepsilon^{-j(z+z'-2d)}}{(z + z' - 2d)(-j + n^{-1}\tan\beta nd)} \quad (9$$

where there are no residue terms since $\beta z \gg 1$ and θ_1 $\pi/2 >$ arc cos $\beta/\lambda_J[A_J(\theta_1) = 0]$.

* See Appendix 13.2 for the exact condition.

t is seen that $I\Big|_{r=0} = \lim_{\theta_1=\pi/2} I\Big|$ eqn. (97)

It is therefore a reasonable assumption that eqn. (97) is valid all values of θ_1, $0 \leqslant \theta_1 \leqslant \pi/2$, $\beta R'' \gg 1$.

The integrations of eqns. (30), (31), and (33) are performed in similar manner.

.2) Example of the Application of the Method of B. L. Van der Waerden to the Integration of Eqns. (32) and (34)

As an illustrative example of the application of the method of L. Van der Waerden[13] to the integration of eqns. (32) and (34), nsider the following integral, obtained by applying eqn. (41) eqn. (32).

$$-\frac{1}{4\pi}\int_{W_2}^{-\infty} \frac{(1 - mn^{-2}) \cosh mz' \cosh mz\varepsilon^{-md}\lambda H_0^2(\lambda r)d\lambda}{m(1 \cosh md + mn^{-2} \sinh md)} . \quad (100)$$

Since the integrand of eqn. (100) has two extra branch cuts,) and (4) in Fig. 3, at $\lambda = \pm n\beta$, and one of these intersects e path of steepest descent for the integrand in (100) if the ethod of steepest descents, Appendix 13.1, is used, the integra- n by the method of steepest descents is complicated. The n der Waerden method contains a systematic procedure for tegrating around branch cuts.

In order to put eqn. (100) in Van der Waerden's form the llowing transformation is performed:

$$\lambda = -j\mu \quad . \quad . \quad . \quad . \quad . \quad . \quad (101)$$

The asymptotic representation for large λr is made for $H_0^2(\lambda r)$. The transformed $\mu = j\lambda$-plane is shown in Fig. 23, where), (2), (3), (4) and (5) are the branch cuts that appear in the plane, Fig. 3.

Then $$I = \int_{W_2} W\varepsilon^{-ru}dv \text{ for } \beta r \gg \beta d \quad . \quad . \quad . \quad (102)$$

here r corresponds to Van der Waerden's λ.

In the μ-plane the saddle-point of the α-plane becomes nother branch point.[13] In this case there is no added com- ication because this added branch point is coincident with the e at $\mu = j\beta$.

For large r and for a contour shifted far to the right in Fig. 23, e integral reduces to integrations over the branch cuts and oles. The former yield rapidly converging series for large λr d for poles not too close to the branch points.

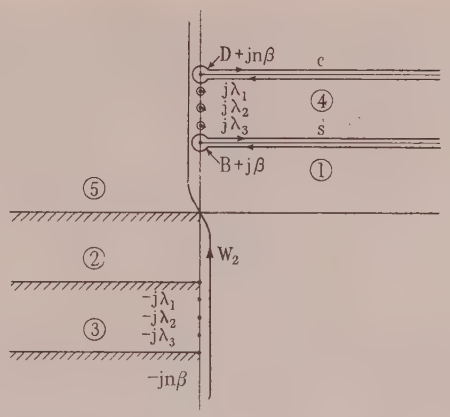


Fig. 23.—Paths of integration for cases B-2, C-2 in the $u = j\lambda$-plane.

Performing the integrations exactly as Van der Waerden does, the following is obtained:

$$I = -2\pi j \sum_{j=1}^{k} \text{Residue} - \frac{1}{2\pi r}\varepsilon^{-jn\beta r} + 0(r^{-2}) \quad . \quad (103)$$

where $0(r^{-2})$ indicates a term of order r^{-2} and yields an estimate of the error.

The residues are computed in the usual way and are modified by eqn. (42) to yield the summation terms of eqn. (48).

If in eqn. (32) it is assumed that

$$\left.\begin{matrix}\beta z' \\ \text{and} \\ \beta z\end{matrix}\right\} \leqslant \beta d \ll \beta r$$

then $\beta R^V \simeq \beta R' \simeq \beta r$ and eqn. (48) is obtained.

It has been assumed that the poles are not too near to the branch points. An estimate of what is not too near is a result of the Van der Waerden method. It is

$$\left.\begin{matrix}\varepsilon^{-r|\beta-\lambda_1|} \\ \text{and} \\ \varepsilon^{-r|n\beta-\lambda_k|}\end{matrix}\right\} \ll 0(r^{-2}) \ll 0(r^{-1}) \quad . \quad . \quad . \quad . \quad (104a)$$

$$. \quad . \quad . \quad . \quad (104b)$$

If the integrations in Appendix 13.1 had been performed by the method in Appendix 13.2, the definition of the pole being not too near the branch point would be found to be (104*a*). Hence eqn. (104*a*) and (104*b*) are necessary restrictions for all the eqns. (45)–(50).

621.396.11.029.4

Monograph No. 114
Jan. 19

THE IONOSPHERIC PROPAGATION OF RADIO WAVES OF FREQUENCY 16KC OVER SHORT DISTANCES

By T. W. STRAKER, M.Sc., Ph.D.

(*The paper was first received 29th September*, 1953, *and in revised form 21st September*, 1954. *It was published as an* INSTITUTION MONOGRA *in January*, 1955.)

SUMMARY

The paper describes a continuation of the investigations of Best, Budden, Ratcliffe and Wilkes on the ionospheric propagation of radio waves of frequency 16kc/s emitted from the Post Office sender GBR at Rugby, and observed at Cambridge, 90km away. The present experiments over the period March, 1948, to October, 1949, have confirmed the tentative conclusions of the earlier work, and revealed some new and interesting effects.

Most of the observations were made on the abnormal component of the downcoming wave, using the method previously developed to isolate the downcoming wave and to make separate measurements on its amplitude and phase. On a few occasions simultaneous observations were made on two frequencies transmitted over comparable distances.

The observations showed that reflection took place from a height of about 72km by day and 87km at night, and that the daily cycle of phase change was related to a real change in the height of reflection. The apparent height of reflection usually varied in a regular and predictable way with the zenith angle of the sun. The total change of height in passing from midday to midnight was 16·8km in summer and 13·5km in winter. There was a marked seasonal variation of the amplitude of the downcoming wave, the variation of the day-time amplitude being different from that of the night-time amplitude. The conversion coefficient was 0·13 on a summer day and 0·26 on a winter day, whereas at night it was 0·56 in summer and 0·37 in winter. During the summer months the amplitude was closely controlled by the sun, but it was not related in any simple way to the zenith angle. The daily cycles of amplitude change and phase change appeared to take place independently. The sunrise effect on the amplitude occurred about one hour before the effect on the phase, the latter occurring about 9min after ground sunrise.

Some subsidiary variations of the phase and amplitude, which were regularly observed, were explicable on the assumption that reflection took place from two levels in the upper atmosphere. There was no clear evidence of waves reflected twice from the ionosphere.

During and after great magnetic storms the diurnal variation of the height of reflection was markedly abnormal, the abnormality lasting for as long as 10 days after the end of the magnetic disturbance. Anomalous behaviour of a less violent nature was frequently observed both by day and by night. Some of these anomalies could be directly associated with magnetic disturbances, but in general there was no simple connection between the two phenomena.

(1) INTRODUCTION

Since 1934 part of the programme of radio research at the Cavendish Laboratory has been directed towards an understanding of the ionospheric propagation of low-frequency waves. The pre-war work, described in two papers by Best, Ratcliffe and Wilkes,[1] and Budden, Ratcliffe and Wilkes,[2] included a long series of measurements on waves of frequency 16kc/s observed at a fixed distance of 90km. After the war this work was extended to greater distances and to higher frequencies, and the results have been described in a series of companion papers.[3–9] As part of the post-war programme it was considered desirable to make a longer series of detailed observations on 16kc/s at distance of 90km, both to test the tentative conclusions of t pre-war work and for comparisons with the measurements higher frequencies and over greater distances. The purpose the paper is to describe these measurements on a frequency 16kc/s and to give an account of the results obtained.

No attempt will be made here to relate the observations those obtained on higher frequencies and over greater distanc This has been done in the companion papers and also in summarizing paper by Bracewell, Budden, Ratcliffe, Strak and Weekes.[10] Where it is appropriate, however, the resu will be compared with those obtained before the war for t same frequency and the same path.

In order to study the characteristic properties of the dow coming wave at short distances from the sender, Best, Ratcli and Wilkes[1] developed a simple method by which changes the phase and amplitude could be separately measured. T method has been used in the present investigations. A br outline of the method, together with a description of expe mental arrangements using a new technique, is given in Section The analysis and interpretation of the results, and a way of ov coming the difficult problem of suppressing the ground wa are also discussed in the same Section. The results are d cussed in Sections 3–7. In Section 8 an attempt is made relate the anomalous propagation effects, sometimes observe to disturbances in the magnetic field of the earth.

In Section 3 measurements are described which lead to determination of the apparent height of reflection by using adaptation of the frequency-change method of Appleton a Barnett.[11]

(2) THE PRESENT EXPERIMENTS

In the present experiments a long series of measurements w made on signals of frequency 16kc/s emitted by the Post Off sender, GBR, at Rugby, and received at Cambridge, 90km awa using a method previously developed[1] for measuring the pha and amplitude of the downcoming wave relative to the grou wave. From these measurements it was then possible to estima the reflection coefficient of the ionosphere and changes in t apparent height of reflection of the downcoming wave.

In a second kind of experiment, simultaneous measuremer of the kind described above were made over short periods signals from GBR and signals of frequency 19·4kc/s emitted the Post Office sender, GBZ, Criggion, at a distance of 222k from Cambridge. From these measurements it was possible estimate the apparent height of reflection.

(2.1) Experimental Method

The experimental method has been described elsewhere[1,5] some detail and a brief outline only will be given here.

The electromagnetic field at the receiver may be regarded being composed of a ground wave with electric and magne fields E_0 and H_0, and a downcoming wave incident at an angle In general, the downcoming wave is elliptically polarized a

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Straker is with the Defence Research Board, Canada.

ay be resolved into two plane-polarized components, the rmal component (E_1, H_1) with its magnetic field perpendicular the plane of propagation,* and the abnormal component $_2$, H_2) with its magnetic field in this plane.

The experimental method consists in isolating the abnormal mponent by using a loop aerial with its plane perpendicular the plane of propagation so that it receives an e.m.f. proporonal to $2H_2 \cos i$, and measuring its amplitude and phase lative to the ground wave. In the present experiments the n.f. induced in a vertical aerial was used as the phase reference. his e.m.f. is proportional to $H_0 + 2H_1 \sin i$, and it was verified at its phase did not differ from that of the ground wave by ore than 7°.

(2.2) Experimental Arrangements

The experimental arrangements may be understood by ference to Fig. 1. The e.m.f. induced in a small loop aerial,

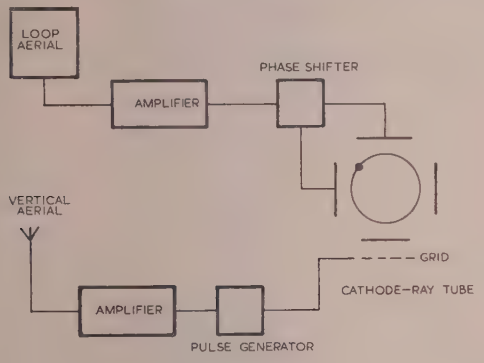


Fig. 1.—The experimental arrangements for visual observations.

ented to receive the abnormal component alone, was amplified id suitably divided and phased to produce a circular time-base a cathode-ray oscillograph. The reference e.m.f. induced in short vertical aerial was amplified and used to generate pulses length $\frac{1}{5}$ microsec, one each radio-frequency cycle. These ulses were applied to the grid of the cathode-ray tube to modulate e intensity of the circular time-base. The polar co-ordinates the well-defined bright spot obtained in this way completely presented the amplitude of the abnormal component and its ase relative to the reference e.m.f. For visual observation, e intensity of the illumination on the cathode-ray tube was djusted so that only the bright spot was visible; its position was en read on a transparent polar scale fixed to the face of the be.

In the method used for automatic observations the Cartesian o-ordinates of the position of the bright spot were recorded as a nction of time on moving photographic paper. This was hieved by removing the y-deflecting voltage from the cathodey tube and shifting the phase of the reference e.m.f. by 90° uring alternate half-minutes. A brightening voltage was applied the grid of the tube at intervals of 6 min to provide additional ming marks. Typical examples of the photographic records btained in this way are shown in Fig. 2. Examples (*a*) and (*c*) re typical of the records obtained by day and by night, while xample (*b*) shows the effects of sudden anomalies of the type ssociated with solar flares, at 1439h, and 1532h, G.M.T. It interesting to note that an accurate timing reference was pplied by the nature of the time signals from GBR at 1000 nd 1800h, G.M.T. From the photographic records it was

* The plane of propagation is defined as the vertical plane containing the sender nd the receiver.

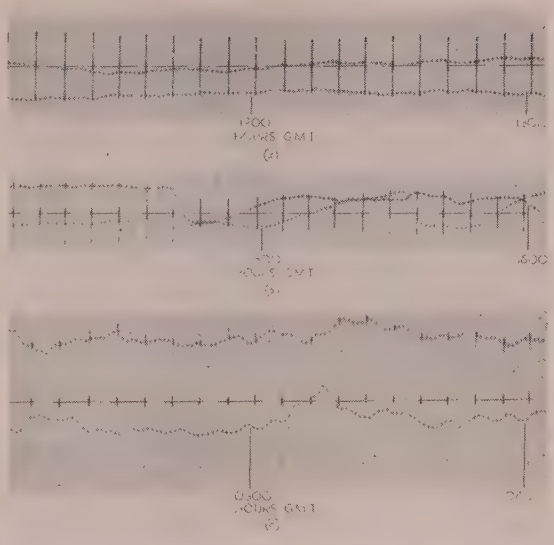


Fig. 2.—Typical examples of the records obtained with the automatic equipment.
(*a*) 14th October, 1948.
(*b*) 6th October, 1948.
(*c*) 11th October, 1948.

possible to estimate the time to an accuracy of 1 min, the phase to 3°, and the amplitude to 5%.

(2.3) Suppression of the Ground Wave

The characteristics of the downcoming wave are deduced in later Sections on the assumption that the loop aerial has been accurately set to receive the abnormal component and suppress the much stronger ground wave. If there were no downcoming wave at any time during the day, no uncertainty would arise in setting a loop to suppress the ground wave. However, a strong downcoming wave is always observed at very low frequencies, and it is difficult to know when the ground wave has been completely suppressed. A method for doing this by trial and error has been proposed by Best, Ratcliffe and Wilkes.[1] As this method is based on an important assumption regarding the behaviour of the downcoming wave, it is proposed to examine it in the light of the present experiments.

In the first instance the loop was set as nearly as possible with its plane perpendicular to the plane of propagation, and observations of the phase and amplitude of the received e.m.f. were made throughout the day and a polar curve of the type shown in Fig. 3 was plotted. Almost invariably, it was found that the polar curve was circular in the period from sunrise to sunset and that the origin was within the circle but not at its centre. It was assumed that some e.m.f. induced by the ground wave caused the centre to be displaced from the origin, and on subsequent days the loop setting was adjusted until the centre of the polar curve lay as near as possible at the origin. This procedure involves the important assumption that during the daylight hours the abnormal component changes in phase while its amplitude remains substantially constant. Evidence which strongly supports this assumption is given in the next paragraph.

For the long series of observations on signals from GBR the loop setting was initially adjusted in the manner described above, and afterwards it was left unchanged during systematic observations over a period of 18 months. The curves of Fig. 3 are typical of the polar curves obtained at fortnightly intervals by plotting the phase and amplitude averaged over 14 days.* In

* The method of averaging is explained in Section 2.4.

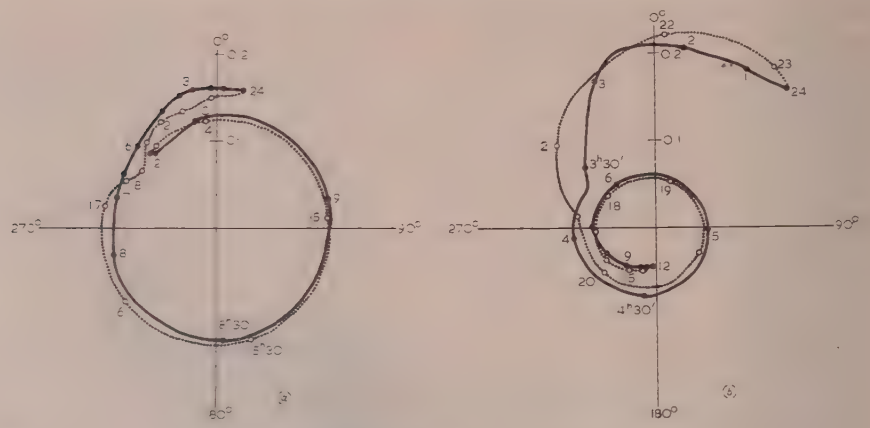


Fig. 3.—Polar curves obtained by plotting the phase and amplitude of the abnormal component.

(*a*) Results averaged over the period 1st to 15th December, 1948.
(*b*) Results averaged over the period 15th to 30th June, 1949.

all cases the polar curves during the daylight hours were approximately circles with their centres very nearly at the origin. The simplest explanation of these results is that the amplitude of the abnormal component remains substantially constant while its phase changes.

(2.4) Analysis and Interpretation of the Results

Observations on signals from GBR were made at Cambridge on 362 days* between 11th March, 1948, and 30th September, 1949. On 266 days the observations extended over the periods from sunrise to midday or from midday to sunset, and sometimes over both periods, while on 109 occasions the observations extended throughout the night. Measurements of the phase and amplitude of the abnormal component were obtained from the automatic records at intervals of 6 min, and the measurements were plotted against the time of day in the way shown in Fig. 4.

In order to determine the regular behaviour of the downcoming wave, it was necessary to eliminate the effects of disturbed periods, and also to use some process of averaging to eliminate the small variations from day to day. The daily plots such as those in Fig. 4 showed that on most days the phase and amplitude changed smoothly and regularly, whereas on some days they were more or less violently disturbed. These disturbed periods† were usually characterized by large changes of phase which were obvious from an inspection of the plots, and the measurements were excluded from the analysis of the regular behaviour. If the curves for each morning (or afternoon) in any 14-day period were superimposed so that the changing times of sunrise (or sunset) coincided, they were found to be closely similar. The curves were therefore grouped in sets corresponding to 14 days, and mean curves were drawn through them. These mean results were used to investigate the regular behaviour of the downcoming wave.

The results will be described in terms of a model in which the ionospheric wave is reflected, possibly with a change of phase, from a sharply-defined reflecting surface with a change of amplitude measured by a reflection coefficient. It is not intended to imply that the ionospheric wave is, in fact, returned to earth in this way. The height of this fictitious reflector above the surface of the earth will be referred to as the "apparent height

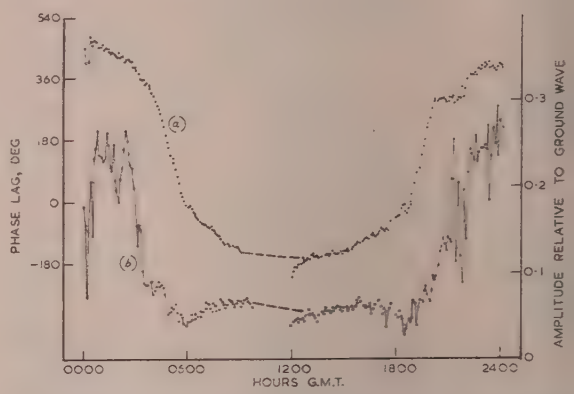


Fig. 4.—Phase and amplitude of the abnormal component dedu from the automatic records, 3rd July, 1948.

(*a*) Phase curve.
(*b*) Amplitude curve.

of reflection," and in the model, changes in the phase of downcoming wave will be described as if they were due entir to changes in the apparent height of reflection.

As the wave incident on the ionosphere (E^i, H^i) is linea polarized with its electric vector in the plane of propagati whereas the reflected wave is, in general, elliptically polariz it is convenient to distinguish between the reflection coefficie for the normal component (E_1^r, H_1^r)* and the abnormal compon (E_2^r, H_2^r).† The ratio H_1^r/H^i will be referred to as the reflect coefficient $_{||}R_{||}$, and the ratio H_2^r/H^i as the conversion coe cient $_{||}R_{\perp}$. The way in which these coefficients are calcula from the amplitudes of the e.m.f.'s received in the aerials been explained in detail elsewhere[5,8] and will not be repea here.

In most of the experiments described in the paper observati were made on the abnormal component alone, and it was assur that its behaviour represented that of the downcoming wave a whole. This assumption is supported by extensive measureme on very-low-frequency waves observed over short distances,[1] which showed that the state of polarization of the downcom

* After the end of 1948, transmissions from GBR were somewhat irregular, and it became necessary to ask for special transmissions, both to ensure the continuity of the measurements and for comparison with other observations.

† These disturbances are discussed in some detail in a later Section.

* A symbol with a superscript *r* represents the magnitude of the component im diately after reflection from the fictitious reflector.

† The abnormal component is polarized with its electric vector perpendicula the plane of propagation, the normal component with its electric vector in this pl

ve was always approximately constant. In addition, as the io H_1^r/H_2^r was always near to unity, it seems reasonable to ume that the reflection coefficient $_{||}R_{||}$ is roughly equal to conversion coefficient $_{||}R_{\perp}$.

(3) THE APPARENT HEIGHT OF REFLECTION

Although there is a fairly detailed knowledge of diurnal and sonal variations of the phase and amplitude of very-low-quency waves reflected at nearly vertical incidence from the osphere, precise information regarding the levels in the upper nosphere to which this knowledge applies is much less com-te. At the low frequencies concerned it is not possible to use pulse technique of Breit and Tuve[12] to determine the height reflection, and indirect methods must be used. In this Section

where h = Apparent height of reflection.
R = Radius of the earth (taken as 6 368 km).

In the experiments described below the value of the integer $(a_1 - a_2)$ is taken as unity, no other value giving a reasonable solution for the height of reflection. A measurement of $(\phi_1 - \phi_2)$ then enables the apparent height to be determined.

On 4th and 5th October, 1949, simultaneous observations were made at Cambridge on signals from the senders GBR (16 kc/s, 90 km) and GBZ (19·4 kc/s, 222 km). The phases and amplitudes of the abnormal components of the two signals were recorded automatically in the way described in Section 2.2, using two separate equipments with a common timing unit. The phase results for 5th October are shown in Fig. 5(*a*); similar curves were obtained on 4th October. It will be noticed that

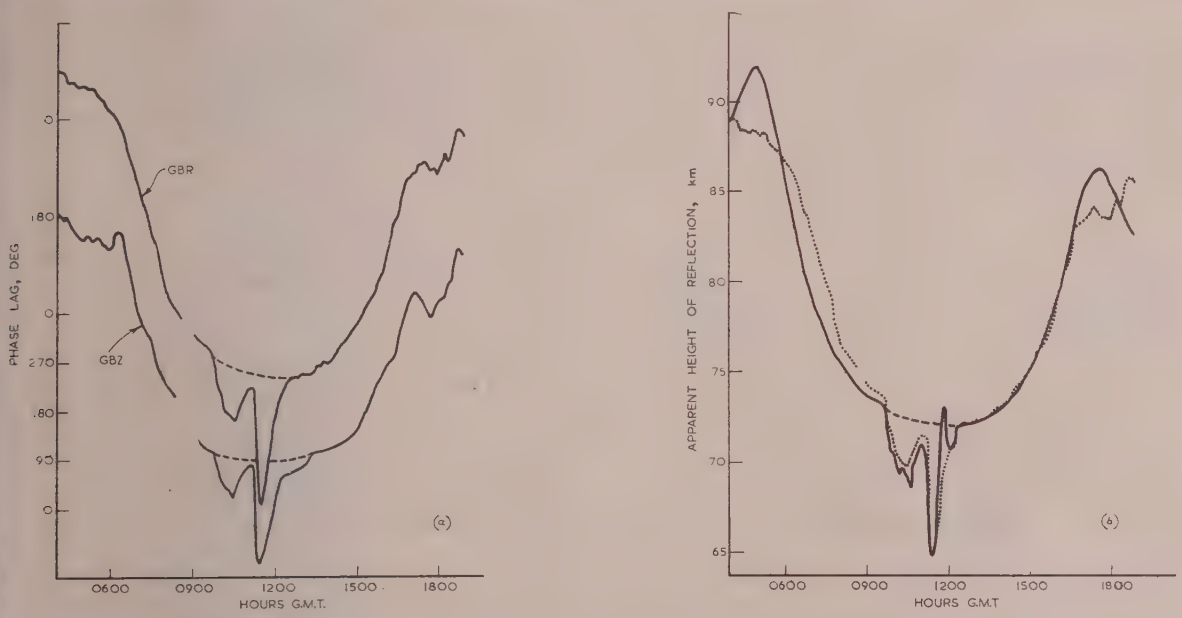


Fig. 5.—Determination of the height of reflection by the frequency-change method, 5th October, 1949.

(*a*) Observed phase of the abnormal component of the downcoming wave from GBR (16 kc/s, 90 km) and GBZ (19·4 kc/s, 222 km).
(*b*) Apparent height of reflection deduced from the phase curves (*a*). The dotted curve shows, for comparison, the variation of reflection height deduced from the phase curve for GBR, on the assumption that changes of phase are entirely due to changes in reflection height.

periments will be described which led to a determination of apparent height of reflection from measurements of phase. The method has been described elsewhere[6] in some detail d only a brief outline will be given here. It consists in measur-the phase difference ϕ between the downcoming wave and ground wave. If it is assumed that waves of nearly equal quencies, f_1 and f_2, propagated over comparable distances, and d_2, are reflected from the same apparent height with the ne change of phase at reflection, then

$$a_1 - a_2 + \phi_1 - \phi_2 = \frac{(s_1 - d_1)}{c} f_1 - \frac{(s_2 - d_2)}{c} f_2 \quad . \quad (1)$$

ere a = Integral number of free-space wavelengths.
c = Velocity of the wave in free space.
s = Equivalent path of the ionospheric wave.
s given with sufficient accuracy by the equation

$$s^2 = 4h^2 + d^2\left(1 + \frac{h}{R}\right)\left(1 - \frac{1}{48}\frac{d^2}{R^2}\right)$$

on 5th October there were two phase anomalies of the type associated with solar flares, one lasting from 0942 h until 1113 h, and the other from 1113 h to 1230 h.

The phase difference between the curves of Fig. 5(*a*) plotted against the time of day shows a mean daily trend together with small periodic oscillations. These oscillations correspond to oscillations in the phase curve for GBZ and are caused by the presence of a wave component which has been reflected twice from the ionosphere. The phase difference $(\phi_1 - \phi_2)$ given by the mean curve was used in conjunction with eqn. (1) to calculate the apparent height of reflection. The results are shown by the full curve in Fig. 5(*b*). It will be noticed that the apparent height varies throughout the day in a similar manner to the phase shown in Fig. 5(*a*). This point is illustrated in Fig. 5(*b*), where the dotted curve which shows the variation of reflection height deduced from the phase curve for GBR, on the assumption that changes of phase are entirely due to changes in reflection height, has been superimposed on the full curve. The same general agreement throughout the daylight hours was found

for the corresponding curves for 4th October. It will also be noticed that there appears to be a definite lowering of the apparent height at the times of the phase anomalies. This conclusion was also confirmed by the results obtained on 4th October.

In Section 4 it is shown that the apparent height of reflection varies with the zenith distance χ of the sun in the simple manner given by eqn. (4). All the observations of apparent height on 4th and 5th October have been summarized in an equation of this form, and lead to the conclusion that $h_{\chi=0} = 69{\cdot}0 \pm 0{\cdot}25$km, where $h_{\chi=0}$ is the apparent height which would be measured if the sun were vertically overhead ($\chi = 0$).

The above results are in close agreement with those obtained by Budden, Ratcliffe and Wilkes,[2] and Weekes,[3] from measurement on the interference pattern produced on the ground by the sender GBR. Their results have been summarized in Reference 10.

(4) CHANGES OF APPARENT HEIGHT OF REFLECTION

On most days the phase of the abnormal component underwent smooth and regular variations which were repeated from day to day. This point is illustrated in a striking way in Fig. 6, which shows the observations plotted at intervals of 12min over a period of 38 days.

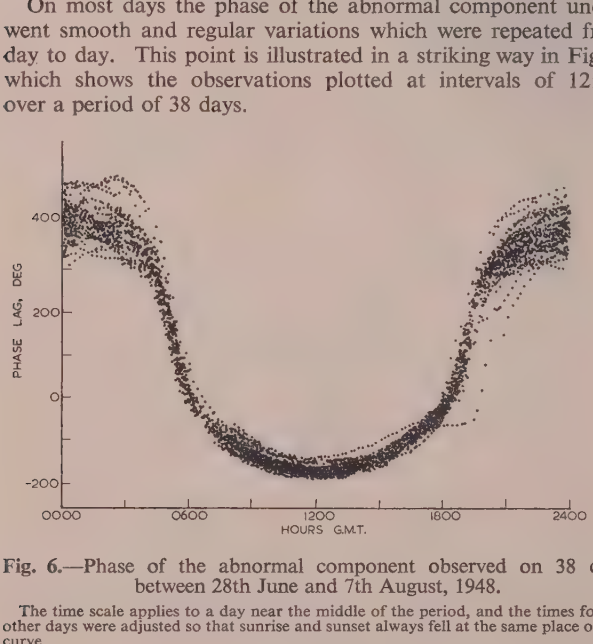


Fig. 6.—Phase of the abnormal component observed on 38 days between 28th June and 7th August, 1948.

The time scale applies to a day near the middle of the period, and the times for the other days were adjusted so that sunrise and sunset always fell at the same place on the curve.

Fig. 7 shows the phase curves, averaged in the way desc in Section 2.4, for the period 3rd April, 1948, to 30th Septe 1949. These curves summarize the regular behaviour o phase of the abnormal component. The times of ground su and sunset at the mid-point between Cambridge and Rugb shown by the fine vertical lines, and the zero of phase for curve is marked on it by a short horizontal line.

As explained in Section 2.4, it is convenient to describ changes of phase on the supposition that they result en from changes of height of a fictitious reflecting surface, sc the curves of Fig. 7 may be regarded as showing the w which the height of reflection varies throughout the day. the conditions of the present experiment a change of pha 100° corresponds to a change of height of 3·0km.

It will be seen from the curves of Fig. 7 that the daily vari of apparent height is closely controlled by the sun. theoretical considerations, Budden, Ratcliffe and Wilkes[2] deduced the following simple expression relating the heig reflection and the zenith distance of the sun, on the assum that reflection took place from a level of constant ioniz density in the lower parts of an ionized region of the type cussed by Chapman:[13]

$$h_\chi = h_{\chi=0} + h_s \log \sec \chi \quad . \quad . \quad . \quad .$$

where χ = Zenith distance of the sun.
h_χ = Apparent height when the sun's zenith angle is
$h_{\chi=0}$ = Apparent height when the sun is vertically over ($\chi = 0$).
h_s = Scale height of the atmosphere.

This equation is valid for $\chi < 85°$. If the scale height h_s rer constant, $h_{\chi=0}$ is a constant. If the equation applies t observations, a plot of h_χ against log sec χ should give a str line whose slope gives the magnitude of h_s and whose inte on the line log sec $\chi = 0$ gives the magnitude of $h_{\chi=0}$.

The accuracy with which the results fit eqn. (2) has tested by plotting h_χ against log sec χ. The results fo morning of 3rd July, 1948, are shown in Fig. 8. On most the results could be represented by a straight line as close in Fig. 8, while during winter days the fit was much closer.

It will be noticed in Fig. 8 that the plotted points depart the straight line in a regular manner; a similar behaviour noticed in all the plots from late February to early Oct but it was not observed during the other (winter) months

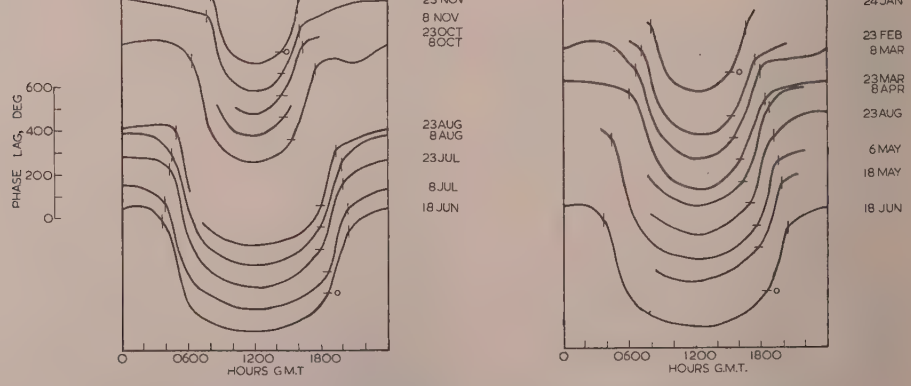
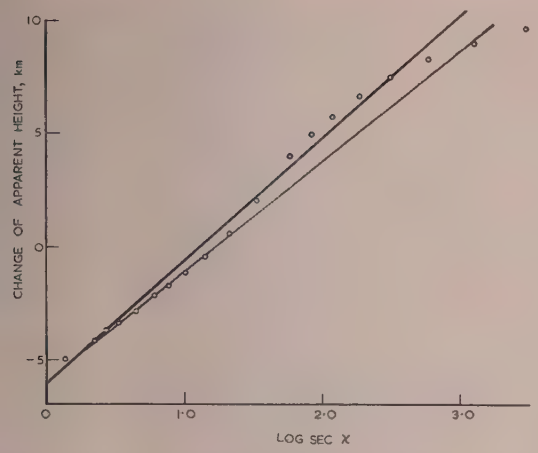


Fig. 7.—Diurnal variation of the phase of the abnormal component in 1948 and 1949.

[' Time of sunrise and sunset.
— ' Zero phase relative to ground wave.



ig. 8.—Variation of apparent height of reflection with log sec χ from sunrise to midday, 3rd July, 1948.
- - - - Results obtained from midday to sunset on the same day.

ction 7 reasons are given for supposing that these small riations are produced by a subsidiary wave component reflected om a lower level in the ionosphere so as to interfere with the ain component. In the light of this supposition, it is con-ered that the curves relating the height of reflection of the ain abnormal component to log sec χ do not depart from aight lines by more than $\pm 0{\cdot}3$km (10° in phase).

The dotted line in Fig. 8 refers to the results obtained in the ernoon of 3rd July, 1948, and it will be seen that the slope of e line for the morning is markedly greater than that for the ernoon. This type of asymmetry was always present, and its gnitude was independent of the time of year. The mean ference between the slopes for the mornings and the afternoons the whole period of the observations was 28° (0·84km). is asymmetry could be produced if the rate of disappearance electrons was small enough.

From the average curves of the type shown in Fig. 7, straight es of the type shown in Fig. 8 have been plotted and their pes and intercepts determined for different times of the year. e results are shown in Fig. 9. The values obtained by averaging intercepts for the morning and the afternoon are plotted ainst the time of year in Fig. 9(*b*). There appears to be a all seasonal variation of $\pm 0{\cdot}5$km, with a maximum in the mmer and a minimum in the winter. To a first approximation, wever, the intercept is constant, implying that $h_{\chi=0}$ in eqn. (2) constant.

The values obtained by averaging the morning and afternoon pes are plotted against the time of year in Fig. 9(*a*). It will noticed that there is a marked semi-annual variation. A urier analysis of the data showed that the slope (S) could be pressed in the form

$$S = 5{\cdot}52 + 0{\cdot}43 \sin (n - 78)^\circ + 1{\cdot}02 \sin (2n - 88)^\circ \quad (3)$$

ere S is in kilometres and n is the number of days from January expressed in degrees. This equation has been used calculate the full curve in Fig. 9(*a*). The maxima in the semi-nual term occur on 29th March and 29th September, very arly at the equinoxes; the maximum and minimum in the nual term occur on 18th June and 19th December, very nearly the summer and winter solstices.

The fact that the results agree roughly with eqn. (2) cannot mediately be taken to imply that the apparent height corre-onds to a height of constant electron density. Only a full-

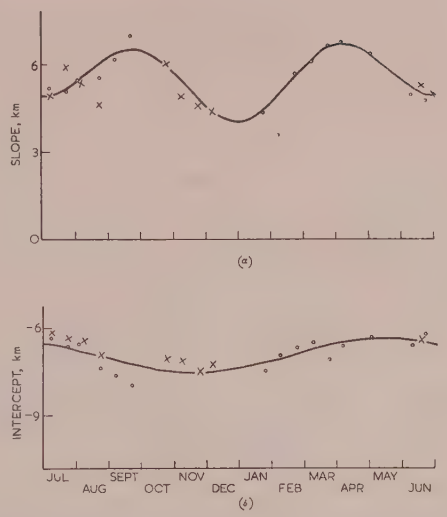


Fig. 9.—Seasonal variation of (*a*) the slopes and (*b*) the intercepts of the straight lines, such as those shown in Fig. 8.
× Observations made in 1948.
○ Observations made in 1949.

wave theory of the reflection of very long waves can settle the point, and such a theory does not seem to be available at present. It is, however, interesting to consider what deductions could be made if the apparent height of the fictitious reflector were coincident with a level of constant electron density in the tail of a Chapman layer. Eqn. (2) then shows that the slope of the straight line obtained when log sec χ is plotted against h_χ would give the magnitude of the scale height h_s. On this supposition the scale height is found to be 5·5km on the average. This is near the value of 6km found by Budden, Ratcliffe and Wilkes,[2] and, as they pointed out, corresponds to a temperature of 180°K on the supposition that the air has the same composition as at the ground. Recent experiments with rockets by Havens, Koll and LaGow[14] have confirmed that the temperature is of this order at these heights. There does not seem to be any independent evidence that the temperature or composition of the atmosphere at these heights changes in any way which would account for the seasonal change of the slopes of the lines discussed in this Section.

Summarizing, it may be said that the apparent height of reflection may be represented by an equation of the form

$$h_\chi = h_{\chi=0} + A(t) \log \sec \chi \quad . \quad . \quad . \quad . \quad (4)$$

where the symbols have the same meaning as in eqn. (2), and $A(t)$ is a function given by eqn. (3), or by the curve of Fig. 9(*a*). The results indicate that $h_{\chi=0}$ is approximately constant with a value as suggested in Section 3 of 69km. It should be emphasized that the above expression does not apply to higher frequencies[8] and greater distances.[6]

The phase of the abnormal component at local noon and at night (averaged over the period from 2100 to 0300h, G.M.T.) is plotted against the time of year in Fig. 10. The scale for the apparent height has been fitted to the results on the assumption that $h_{\chi=0} = 69$km as suggested in Section 3. The seasonal variation of the height at local noon arises from the variation of the factor $A(t)$ log sec χ in eqn. (4), and the full curve shown in Fig. 10 has been calculated from this expression. The results for the height at night are of particular importance as they apply to conditions when the influence of solar radiation is absent. Although the results are not as numerous as those

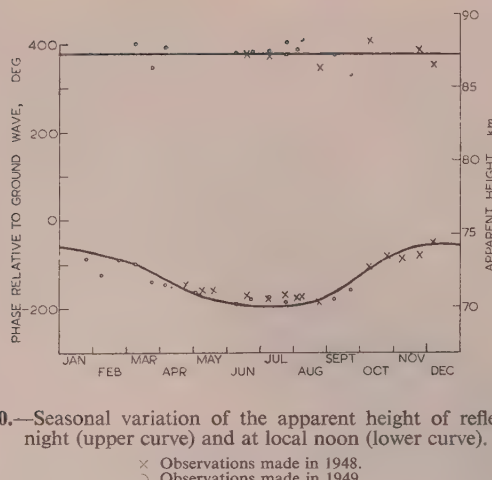


Fig. 10.—Seasonal variation of the apparent height of reflection at night (upper curve) and at local noon (lower curve).

× Observations made in 1948.
) Observations made in 1949.

obtained by day, it seems possible to say that any seasonal variation, if present, is not of large magnitude. In particular, there is no large semi-annual variation of the type found in the daytime. The results of observations at night are explicable on the assumption that reflection takes place from a height of about 87km at all times of the year.

The results described in this Section are consistent with those obtained with the same sender and over the same path in 1938 and 1939.[2] The asymmetrical diurnal variation of the height was not noticed by the pre-war workers, since most of their observations were restricted to the afternoons. The semi-annual variation of the height h_χ was obscured in their results by phenomena associated with great magnetic storms.

(5) THE AMPLITUDE OF THE DOWNCOMING WAVE

On most days the amplitude of the abnormal component underwent smooth and regular variations. These were closely repeated from day to day, but the agreement was not as close as with the measurements of phase. This point is illustrated in Fig. 11, which shows the observations plotted at intervals of

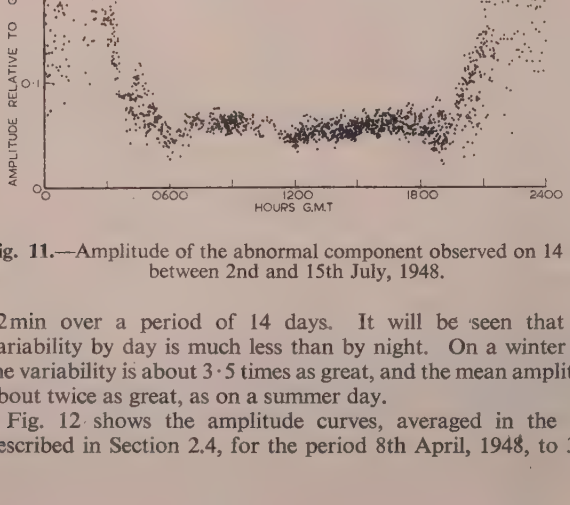


Fig. 11.—Amplitude of the abnormal component observed on 14 days between 2nd and 15th July, 1948.

12min over a period of 14 days. It will be seen that the variability by day is much less than by night. On a winter day the variability is about 3·5 times as great, and the mean amplitude about twice as great, as on a summer day.

Fig. 12 shows the amplitude curves, averaged in the way described in Section 2.4, for the period 8th April, 1948, to 30th September, 1949. The most striking feature of these curve that the amplitude of the downcoming wave is at all times fa high. In winter the amplitude is substantially constant through the day and night. From March until October the behaviou markedly different. About one hour before sunrise at ground the amplitude decreases quite rapidly, remains nearly constant throughout the daylight hours and, after gro sunset, increases rapidly to the night time value. Althoug summer the amplitude is closely controlled by the sun, it is related in any simple way to the zenith distance of the sun to the phase of the downcoming wave. It appears that region of ionization responsible for the absorption of ionospheric wave varies differently from that which ca reflection.

From an inspection of all the records it was clear that t was no appreciable asymmetry in the daily variation of amplitude, of the type found for the phase variations. H again it appears that the reflecting and absorbing regions differently.

Fig. 13 shows the conversion coefficients by day and by n plotted against the time of year. The average amplitude of abnormal component from one hour after sunrise to one h before sunset has been used in making the daytime plot, and average amplitude from one hour after sunset to one h before sunrise has been used at night. The most notice feature of the daytime curve is the rapid increase in amplit in October, followed by a gradual decrease in the spring. inspection of the day-to-day records shows that the main of the transition from summer to winter conditions occu in the course of about four days centred on 18th October. sudden change is similar to the "November effect" reported the early workers,[15–17] who noticed large changes in the stren of low-frequency waves received across the Atlantic, withi period of a few days in November.

It will be noticed in Fig. 13 that the conversion coefficient night undergoes a marked seasonal variation. Here again results for the amplitude differ from those for the phase.

No simple explanation can be offered for the diurnal seasonal variations of the amplitude. It is abundantly cl however, that the phase and the amplitude vary in an independ manner. This evidence suggests the tentative hypothesis reflection and absorption occur at different levels in the up atmosphere.

(6) PHENOMENA OCCURRING NEAR SUNRISE

The curves of Figs. 7 and 12 show that the daily cycles phase change and amplitude change usually start quite sudde near sunrise, and that the times of these phenomena are clos controlled by the sun. From March to October there wa sudden decrease of amplitude occurring about one hour bef ground sunrise, whereas during the other (winter) month sunrise effect of this type was not clearly evident. No co spondingly large changes of phase were observed before grou sunrise. The reverse changes in the phase and the amplit were observed near ground sunset. It will be noticed that daily phase variation took place while the amplitude remai substantially constant.

Fig. 14 shows the times of occurrence of the sunrise effects the phase and the amplitude plotted against the time of y In all cases the points were obtained by drawing two strai lines to fit the average daily curves, one before and one a the sudden change, and taking the time at which the lines in sected. The sudden change in the phase occurred, on the avera 9min after ground sunrise at the mid-point between the sen and receiver, and the sudden change of amplitude from Ma to October occurred 56min before ground sunrise. The

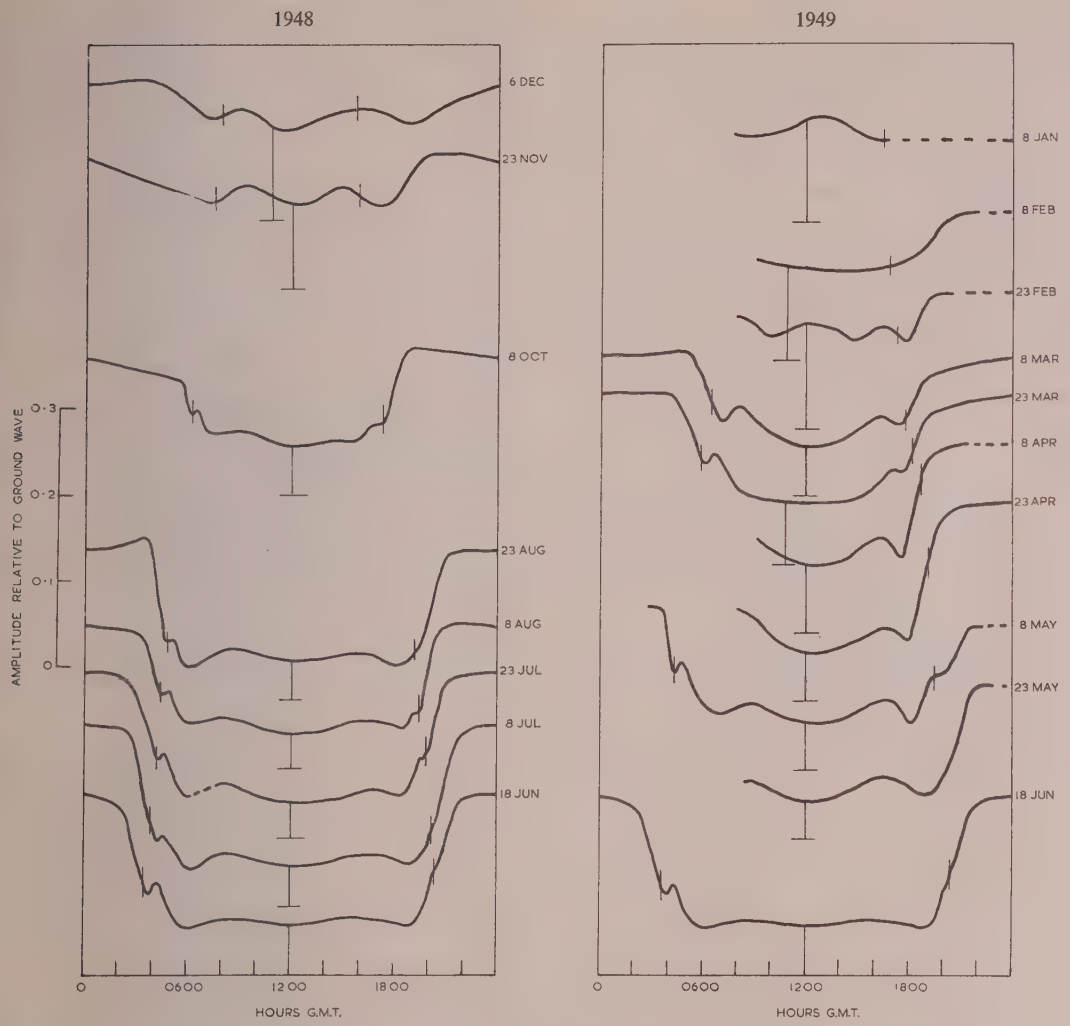


Fig. 12.—Diurnal variation of the amplitude of the abnormal component in 1948 and 1949.

Times of sunrise and sunset.

rve in Fig. 14 shows the time 9 min after ground sunrise. The tted curve shows the time 56 min before sunrise.

(7) POSSIBILITY OF MORE THAN ONE DOWNCOMING WAVE

It will be noticed that there are small but regular variations of e amplitude between sunrise (sunset) and midday in the curves Fig. 12. It has been pointed out in Section 4 that the plots phase against log sec χ revealed the presence of small but gular variations in the diurnal phase change. These variations re present from late February to October, but not during the ner (winter) months. Detailed analysis shows that the small cillations could not have been caused either by a weak twice-lected component or by a weak component of the ground ve, present because the aerial system was not properly adjusted. ere is little doubt that both phase and amplitude variations e caused by a small wave component which interferes with the in reflected component. This weak component might be lected from the region responsible for absorption of the main

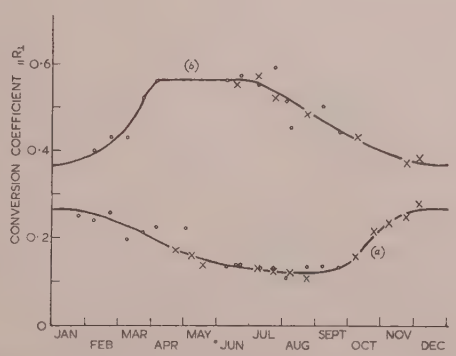


Fig. 13.—Seasonal variation of the conversion coefficient $_{\parallel}R_{\perp}$ by day [curve (*a*)] and by night [curve (*b*)].

× Observations made in 1948.
○ Observations made in 1949.

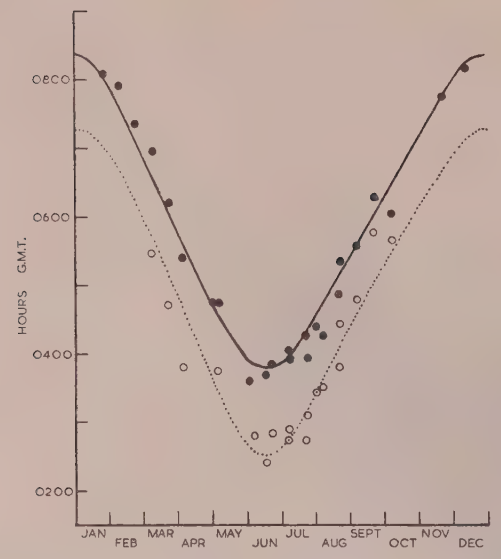


Fig. 14.—Times of occurrence of the sunrise effects on the phase and amplitude.

——— 9 min after ground sunrise.
– – – – 56 min before sunrise.
● Time at which the phase changes.
○ Time at which the amplitude decreases.

small. These are found to be of two types. The first type irregularity is short-lived, lasts for periods from about t minutes to an hour, and occurs during sudden ionosphe disturbances of the type which are closely associated with sol flares. Observations of this type of anomaly have been report elsewhere[4] in some detail, and will not be discussed in t paper. The second type of anomaly may occur by day or night, may last for periods of several days, and may be associate but not in a very precise way, with disturbances in the magne field of the earth. The most violent and prolonged disturbanc appear to be associated with great magnetic storms. Othe occurring by night or by day may sometimes be directly ass ciated with smaller magnetic disturbances and ionosphe storms of the type which cause irregularities in the propagati of high-frequency radio waves. These types of anomalies w be discussed below.

(8.1) Effects associated with Great Magnetic Storms

In August, 1948, transmissions from GBR were sufficient continuous for the making of detailed observations of the pha and amplitude of the downcoming wave before, during a after the great magnetic storm* of 7th August. In Fig. the upper curve shows the daily variation of the phase f the period 6th to 18th August, 1948, and the lower cur shows the variation of magnetic activity as given by the figures† from the magnetic records taken at Abinger.

This world-wide storm[18] began suddenly at 2300h G.M.T., 7th August, and consisted of two parts of great storm intensi

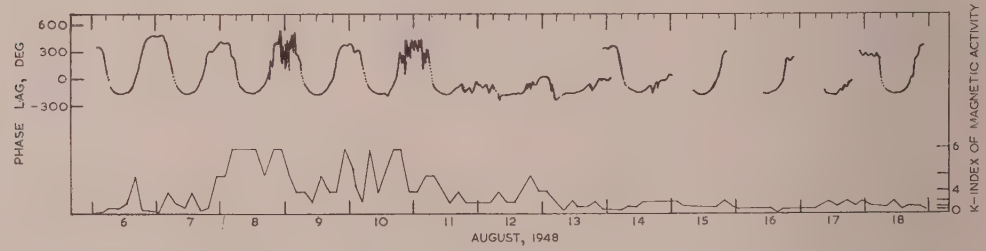


Fig. 15.—Diurnal variation of the phase of the abnormal component and variation of the *K*-index of magnetic activity during and after the great magnetic storm of 7th August, 1948.

component. If it were, the maximum and minimum in the amplitude near sunrise (sunset) would arise from the interference between a partially reflected component which was varying rapidly in phase as the absorbing region was formed, and the main reflected component which was substantially constant in phase at that time. The maximum and minimum amplitudes nearest midday, and also the small variations of phase, would arise from the interference between the main component which was varying in phase and the partially reflected component which, at that time, was varying more slowly. On this supposition the amplitude of the partially reflected component is about one-fifth that of the main component, although the ratio can be as high as one-third.

No clear evidence was found for the presence of a wave reflected twice from the ionosphere, although the results obtained by Bracewell[5] over distances of about 200km indicated that a small component should be observed on a winter day.

(8) ANOMALOUS EFFECTS

If the 14-day average curves for the phase and amplitude are superimposed on a plot showing the individual daily variations, it is possible to detect irregular deviations even if they are quite

from 0300h, 8th August, to 0300h, 9th August, and from 1300 9th August, to 2200h, 10th August. All magnetic observatori agree that the storm had ended by midday, 12th August. T following days in this month, with the exception of the 19 and 20th, were magnetically quiet.

There were no irregularities in the behaviour of the phase the amplitude at the time of the sudden commencement.‡ T most remarkable feature of the observations on the dow coming wave was the absence of irregularities throughout t periods of greatest storm intensity. The first unusual effec were the violent fluctuations of phase and amplitude whi began at 1830h, 8th August, 19 hours after the sudden commenc ment. At times the phase was changing at the rate of 15

* Magnetic storms are classified as "great" or "small" according to the range variation of the magnetic elements.

† The *K* figure refers to the range of magnetic variation during each three-hou period. The lower limits of the range of variation of the horizontal field for th figures obtained at Abinger are

K	0	1	2	3	4	5	6	7	8	9
Range (*r*) ..	0	5	10	20	40	70	120	200	330	500

‡ It is of interest to record that sudden commencements in the Abinger magne records occurred on 16 occasions, some by day and some by night, when observatio on signals from GBR were in progress. There were no irregularities in the behavio of the downcoming wave on any of these occasions.

5km in height) per minute. The behaviour during the follow- day and night was regular, and the records were noticeably e from large fluctuations. Violent fluctuations were observed ain at 1800h, 10th August, about 20 hours after the recurrence intense magnetic activity, and these fluctuations persisted roughout the night.

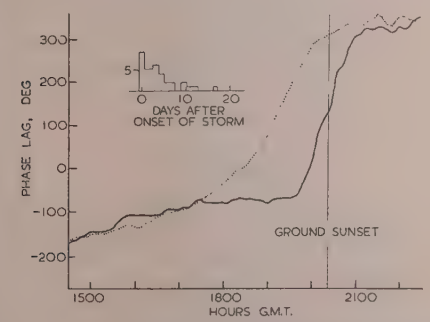
On 11th August, when the magnetic activity had returned to a w level, another and more striking phenomenon was observed the phase variation. Instead of undergoing a regular daily riation, the phase remained substantially constant by day and night until 14th August. Gaps in the transmissions rendered impossible to determine if the height of reflection remained at night-time level or its level at midday. During this period e daily variation of the amplitude was normal. Although on th and 16th August the phase underwent some diurnal varia- on, the slopes of the straight lines obtained by plotting h_χ ainst log sec χ (see Section 4) were markedly less than usual. ormal behaviour was resumed by 18th August. During the ole of the disturbed period the amplitude behaved normally. e accompanying ionospheric storm in the F_2-region was orted as ending at 2300h, 13th August.

Some observations on the downcoming wave were made during e great storm of the 24th January, 1949. Although the results e not as complete as those discussed above, owing to gaps in e transmissions, they indicated the same sort of abnormal haviour for the downcoming wave.

Summarizing, it may be said that there appear to be two stinct types of anomalous behaviour associated with great agnetic storms. The first is characterized by violent fluctua- ons of phase and amplitude, occurring about 20 hours after e onset of great magnetic activity. The second type occurs out three days after the onset of the storm and is charac- rized by an abnormal daily variation of the phase which rsists for as long as 10 days. Similar effects on the ionospheric opagation of low-frequency waves have been observed on gnals of frequency 16kc/s transmitted over a distance of km[2], 16kc/s over 540km,[6] and 57kc/s transmitted across the lantic.[16] Over the longer paths it was found that the anomalous haviour sometimes lasted for several weeks.

(8.2) Disturbances occurring by Day

The example shown in Fig. 16 is typical of a less violent type anomaly which has been frequently observed during the day-



16.—Typical example of anomalous diurnal variation of the phase of the abnormal component, 8th July, 1948.

– – – – Observed phase in the morning, plotted in reflection.

The histogram shows the frequency of occurrence of anomalies plotted against the nber of days after the onset of the nearest magnetic storm.

ne. For the purpose of comparison the curve showing the riation of the phase in the morning has been plotted in reflec- n as a dotted curve. This type of anomaly is mainly charac- terized by an abnormal variation of the phase, such that the plots of h_χ against log sec χ (see Section 4) do not give satisfactory straight lines. From March, 1948, to September, 1949, 40 anomalies were observed during the daytime. With few exceptions the daily variation of the amplitude was normal, and the deviations from the regular variation of the phase represented a decrease of reflection height. Most of the anomalies occurred in the afternoon.

As some of the anomalies could be directly associated with small magnetic storms, an attempt was made to find a general correspondence between the two. The histogram in Fig. 16 shows the frequency of occurrence of anomalies plotted against the number of days after the onset of the nearest magnetic storm. Although there are obvious exceptions, the results encourage the belief that most of the anomalies observed by day are associated with magnetic storms. The apparent tendency for anomalies of this kind to occur about three days after the onset of the storm is interesting in view of the observed delay of three days, previously mentioned in connection with a great magnetic storm. This type of anomaly does not appear to be associated with ionospheric storms in the F_2-layer.

Similar effects have been observed by Bracewell[5] in connection with observations on waves of frequency 15·2kc/s transmitted over a distance of 222km.

(8.3) Disturbances occurring at Night

Apart from smooth variations of about $\pm 30°$ occurring in a period of two or three hours, the phase of the downcoming wave remained substantially constant on most nights. On some occasions, however, large deviations of the phase were observed, and on many of these there was a corresponding decrease of the amplitude. Fig. 17 shows an example of this type of anomaly

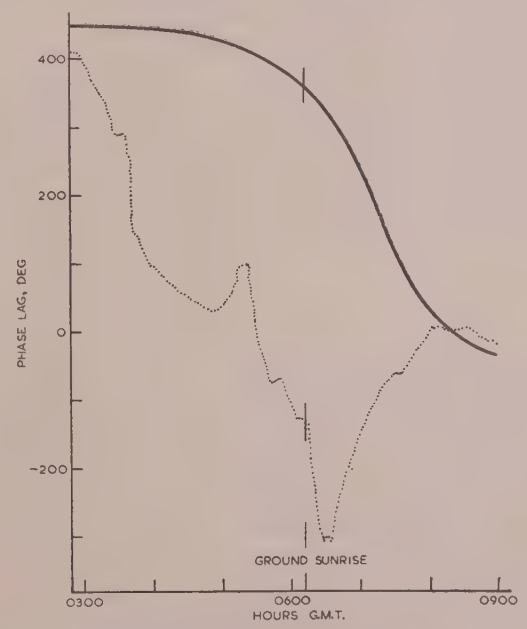


Fig. 17.—Typical example of anomalous variation of the phase of the abnormal component at night, 10th October, 1948.

——— Regular variation obtained by averaging the phase over 14 days.

where the phase deviation was exceptionally large. From March, 1948, to September, 1949, 28 anomalies of this kind were observed at night. With few exceptions, the deviations observed represented a decrease of reflection height. In 10 cases

there was a corresponding decrease of the amplitude by a factor of about 0·5.

Some of the anomalies could be directly associated with small magnetic storms. This point is illustrated in a striking way in Fig. 18, which shows the deviation of the phase from the mean

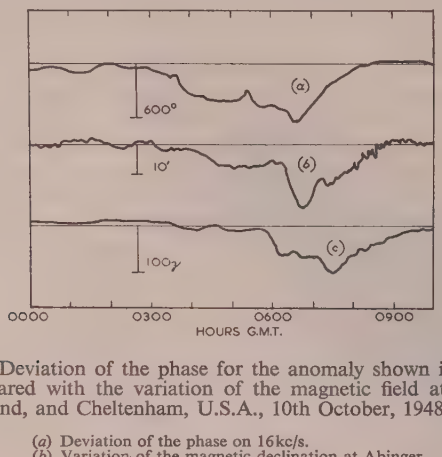


Fig. 18.—Deviation of the phase for the anomaly shown in Fig. 17, compared with the variation of the magnetic field at Abinger, England, and Cheltenham, U.S.A., 10th October, 1948.

(*a*) Deviation of the phase on 16kc/s.
(*b*) Variation of the magnetic declination at Abinger.
(*c*) Variation of the vertical magnetic field at Cheltenham.

curve for the anomaly shown in Fig. 17, together with the variations of the magnetic field at Abinger, England, and Cheltenham, U.S.A. As the phase records and the magnetic conditions were not greatly disturbed on the days preceding and following this disturbance, there can be little doubt regarding the direct connection between the two phenomena. This magnetic storm was world-wide in character, and was accompanied by an ionospheric storm in the F_2-layer.

Anomalies of this type could be clearly associated with magnetic disturbances in 9 cases out of 28. In many cases, however, similar phase effects occurred during magnetically quiet periods, and in other cases there was no phase anomaly when the magnetic elements were disturbed. It is evident that the two effects are not related in a simple manner.

Although ionospheric storms in the F_2-region may be associated with some of these anomalies, it is obvious that the two phenomena are not related in a clear-cut way.

(9) CONCLUSIONS

The principal conclusions may be summarized as follows:

(*a*) Simultaneous observations on two frequencies transmitted over comparable distances lead to the conclusions that reflection would take place from an apparent height of $69{\cdot}0 \pm 0{\cdot}25$km if the sun were overhead, and that the daily cycle of phase change, and changes of phase during sudden ionospheric disturbances, are related to real changes in the height of reflection.

(*b*) The apparent height of reflection varies in a smooth and regular way with the inclination of the sun's rays as given by the equation

$$h_\chi = h_{\chi=0} + A(t) \log \sec \chi \quad . \quad . \quad . \quad . \quad (4)$$

where $A(t)$ is a factor which varies slowly with the time of year with a marked semi-annual periodicity. The height variation is asymmetrically disposed about local noon in such a way that the magnitude of $A(t)$ for the mornings is greater than that for the afternoons. The height of reflection at night is about 87km at all times of the year. The total change of height in passing from midday to midnight is 16·8km in mid-summer and 13·5km in winter.

(*c*) The amplitude of the downcoming wave varies in a regu way throughout the day in summer, while in winter it rema substantially constant by day and night. Marked seaso variations of the amplitude are observed by day and nig the two variations having independent characteristics. T conversion coefficient is about 0·13 on a summer day a 0·26 on a winter day, whereas at night the coefficient is 0·56 summer and 0·37 in winter.

(*d*) The absorption and the height of reflection are clos controlled by the sun, but in different ways. The change absorption occurs about 56min before ground sunrise, where the regular daily changes in the apparent height start abo 9min after sunrise.

(*e*) Small regular variations of the amplitude and the pha superimposed on the main daily variations, are explicable the assumption that reflection occurs from two levels.

(*f*) There are three distinct types of anomalous behavio which appear to be associated with disturbances in the magne field of the earth, but not in a clear-cut way. The anomalo effects which occur during and after great magnetic storms appe to be similar in nature to those observed with low-frequen transmissions over a distance of 540km,[6] and over the Atlantic the anomaly lasting for many days after the end of the magne effect. Anomalies of a less violent nature, which occur by d and by night, appear to be associated with small magne storms, but not in a simple manner.

(10) ACKNOWLEDGMENTS

The work described was carried out as part of the programm of radio research at the Cavendish Laboratory, supported by grant from the Department of Scientific and Industrial Researc Acknowledgment is made to the New Zealand Government, the Council of St. John's College, and to the Director of t Cavendish Laboratory for scholarships and grants whi enabled this work to be carried out.

The author wishes to express his thanks to his colleagues Cambridge who have helped with the experimental observatio particularly to Dr. R. N. Bracewell, Dr. J. Harwood and M P. M. Thompson for helpful discussions and active assistan For assistance with the tedious analysis of the long series records, he is indebted to Mrs. I. Jermyn and Mrs. N. Stanl He also wishes to thank the Post Office Engineering Departme for providing special transmissions from GBR, sometimes short notice. He is greatly indebted to Messrs. H. W. Newt and W. Jackson of the Royal Greenwich Observatory f supplying much information regarding magnetic activity, advance of publication.

(11) REFERENCES

(1) Best, J. E., Ratcliffe, J. A., and Wilkes, M. V.: "Expe mental Investigation of Very Long Waves reflected fr the Ionosphere," *Proceedings of the Royal Society*, 1936, **156**, p. 614.

(2) Budden, K. G., Ratcliffe, J. A., and Wilkes, M. V "Further Investigations of Very Long Waves reflect from the Ionosphere," *ibid.*, 1939, **171**, p. 188.

(3) Weekes, K.: "The Ground Interference Pattern of Ve Low Frequency Waves," *Proceedings I.E.E.*, Paper N 926R, March 1950 (**97** III, p. 100)

(4) Bracewell, R. N., and Straker, T. W.: "The Study Solar Flares by Means of Very Long Radio Waves *Monthly Notices of the Royal Astronomical Socie* 1949, **109**, p. 28.

) BRACEWELL, R. N.: "The Ionospheric Propagation of Radio Waves of Frequency 16kc/s over Distances of about 200km," *Proceedings I.E.E.*, Monograph No. 34R, April, 1952 (**99** IV, p. 217).

) BAIN, W. C., BRACEWELL, R. N., STRAKER, T. W., and WESTCOTT, C. H.: "The Ionospheric Propagation of Radio Waves of Frequency 16kc/s over Distances of about 540km," *Proceedings I.E.E.*, Monograph No. 37R, May, 1952 (**99** IV, p. 250).

) BRACEWELL, R. N., HARWOOD, J., and STRAKER, T. W.: "The Ionospheric Propagation of Radio Waves of Frequency 30–65kc/s over Short Distances," *Proceedings I.E.E.*, Monograph No. 84R, December, 1953 (**101** IV, p. 154).

) WEEKES, K., and STUART, R. D.: "The Ionospheric Propagation of Radio Waves of Frequency near 100kc/s over Short Distances," *Proceedings I.E.E.*, Monograph No. 15R, November, 1951 (**99** IV, p. 29).

) WEEKES, K., and STUART, R. D.: "The Ionospheric Propagation of Radio Waves of Frequency near 100kc/s over Distances up to 1 000km," *Proceedings I.E.E.*, Monograph No. 16R, November, 1951 (**99** IV, p. 38).

) BRACEWELL, R. N., BUDDEN, K. G., RATCLIFFE, J. A., STRAKER, T. W., and WEEKES, K.: "The Ionospheric Propagation of Low and Very Low Frequency Radio Waves over Distances Less than 1 000km," *Proceedings I.E.E.*, Paper No. 1123, May, 1951 (**98** III, p. 221).

(11) APPLETON, E. V., and BARNETT, M. A. F.: "On Some Direct Evidence for Downcoming Atmospheric Reflection of Electric Rays," *Proceedings of the Royal Society*, A, 1925, **109**, p. 621.

(12) BREIT, G., and TUVE, M. A.: "A Test of the Existence of the Conducting Layer," *Physical Review*, 1926, **28**, p. 559.

(13) CHAPMAN, S.: "The Absorption and Dissociative or Ionizing Effect of Monochromatic Radiation in an Atmosphere on a Rotating Earth," *Proceedings of the Physical Society*, 1931, **43**, pp. 26 and 483.

(14) HAVENS, R. J., KOLL, R. T., and LAGOW, H. E.: "The Pressure, Density and Temperature of the Earth's Atmosphere to 160 Kilometers," *Journal of Geophysical Research*, 1952, **57**, p. 59.

(15) ROUND, H. J., ECKERSLEY, T. L., TREMELLEN, K., and LUNNON, F. C.: "Report on Measurements made on Signal Strength at Great Distances during 1922 and 1923 by an Expedition sent to Australia," *Journal I.E.E.*, 1925, **63**, p. 933.

(16) ESPENCHIED, L., ANDERSON, C. N., and BAILEY, A.: "Trans-Atlantic Radio Telephone Transmission," *Proceedings of the Institute of Radio Engineers*, 1926, **14**, p. 7.

(17) HOLLINGWORTH, J.: "The Propagation of Radio Waves," *Journal I.E.E.*, 1926, **64**, p. 579.

(18) "Principal Magnetic Storms," *Terrestrial Magnetism and Atmospheric Electricity*, 1948, **53**, p. 481.

621.396.11 : 538.566.2

Monograph No. 11
Jan. 19

A TABLE OF A FUNCTION USED IN RADIO-PROPAGATION THEORY

By F. HORNER, M.Sc., Associate Member.

(*The paper was first received 17th July, and in revised form 22nd September*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in January*, 1955.)

SUMMARY

Real and imaginary components of the complex integral $1 + 2j\sqrt{(\omega)}\varepsilon^{-\omega}\int_{-j\sqrt{\omega}}^{\infty}\varepsilon^{-x^2}\,dx$ are tabulated as a function of ω. The Tables are prepared for real components of ω in the range 0 to 10, associated with imaginary components in the range -10 to 0.

For the calculation of the field from an aerial, expressions are used which are based on the theory of Sommerfeld[1] and which involve a factor $F(\omega)$ given by[2,3]

$$F(\omega) = 1 + 2j\sqrt{(\omega)}\varepsilon^{-\omega}\int_{-j\sqrt{\omega}}^{\infty}\varepsilon^{-x^2}dx \quad . \quad . \quad . \quad (1)$$

The parameter ω depends on the form and position of the aerial; with an elevated vertical electric dipole, for example,[3]

$$\omega = \frac{-4\pi j(h_t + h_r)^2}{d(1 + \rho_v)^2}$$

where h_t = Height of transmitting dipole above ground level.
h_r = Height of receiving aerial above ground level.
λ = Wavelength.
d = Horizontal distance between transmitter and receiver.
ρ_v = Complex reflection coefficient at the ground for vertically polarized waves.

If the distance d is not small compared with the wavelength, the vertical electric field E can then be expressed in the form

$$E = \frac{60\pi m}{\lambda}\left\{\frac{1}{R_1}\varepsilon^{-2\pi jR_1/\lambda} + [\rho_v + (1 - \rho_v)F(\omega)]\frac{1}{R_2}\varepsilon^{-2\pi jR_2/\lambda}\right\} . \quad (2)$$

in which m = Moment of the transmitting dipole.
R_1 = Length of the direct ray between transmitter and receiver.
R_2 = Length of the ground-reflected ray plane earth.

Eqn. (2) has been given as an example, but $F(\omega)$ and related functions occur more generally in radio-propagation problems.

Norton[4] has given the following series for the evaluation of $F(\omega)$:

$$F(\omega) = -\frac{1}{2\omega} - \frac{1\times 3}{(2\omega)^2} - \frac{1\times 3\times 5}{(2\omega)^3} - \ldots \text{ for } |\omega| > 10 \quad . \quad (3)$$

$$= 1 + j\sqrt{(\pi\omega)}\varepsilon^{-\omega} - 2\omega + \frac{(2\omega)^2}{1\times 3} - \frac{(2\omega)^3}{1\times 3\times 5} + \ldots \text{ for } |\omega| \text{ small} \quad . \quad (4)$$

and has expressed graphically the modulus and phase angle $F(\omega)$, for a wide range of values of ω, in terms of its modul and phase angle. He has also given[2] a short Table of values $|F(\omega)|$, and Burrows[5] has plotted a similar function.

In some types of calculation, in particular those involvi ground reflection coefficients in the form given by McPetrie,[6] is convenient to have values of the real and imaginary parts $F(\omega)$ expressed in terms of the real and imaginary parts of ω.

Tables of values of the function $F(\omega)$ have therefore be prepared by E. J. York at the National Physical Laboratory collaboration with the Radio Research Station. The Ace Pi Model was used for the computations.

Real and imaginary components of ω are denoted by $\mathscr{R}($ and $\mathscr{I}(\omega)$; and three Tables are given. Table 1 is for values $\mathscr{R}(\omega)$ and $-\mathscr{I}(\omega)$ between 0 and 10 at intervals of 0·5. Table is for values of $\mathscr{R}(\omega)$ and $-\mathscr{I}(\omega)$ between 0 and 2 at interv of 0·1. Table 3 is for values of $\mathscr{R}(\omega)$ and $-\mathscr{I}(\omega)$ between and 0·5 at intervals of 0·05.

A Table of an associated function

$$G(\rho) = \sqrt{\left(\frac{\pi}{2}\right)}\varepsilon^{\frac{1}{2}j\pi\rho^2}\int_{\rho}^{\infty}\varepsilon^{-\frac{1}{2}j\pi\lambda^2}d\lambda \quad . \quad . \quad .$$

has been compiled by Clemmow and Mumford[7] for use similar wave-propagation problems. The relationship betwe $F(\omega)$ and $G(\rho)$ is obtained by putting

$$\lambda = x\sqrt{\left(\frac{2}{j\pi}\right)} \quad \text{and} \quad \rho = \sqrt{\left(\frac{2j\omega}{\pi}\right)}$$

and $F(\omega)$ can be obtained from Clemmow's tables by means the relationship

$$F(\omega) = 1 + 2\sqrt{(-j\omega)}G\left[\sqrt{\left(\frac{2j\omega}{\pi}\right)}\right] \quad . \quad . \quad .$$

Clemmow and Mumford effectively tabulate $F(\omega)$ at clos intervals of ω than is done in the present Tables and to t fourth decimal place, but the Tables given here cover a wid range of values of ω.

ACKNOWLEDGMENTS

The work was carried out as part of the programme of t Radio Research Board. This paper is published by permissi of the Director of Radio Research of the Department of Scienti and Industrial Research.

REFERENCES

See page 137.

Correspondence on Monographs is invited for consideration with a view to publication.
The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

Table 1A

Real Part of $F(\omega)$

$-\mathscr{I}(\omega)\rightarrow$

$\mathscr{R}(\omega)\downarrow$	0·0	0·5	1·0	1·5	2·0	2·5	3·0	3·5	4·0	4·5	5·0	5·5	6·0	6·5	7·0	7·5	8·0	8·5	9·0	9·5	10·0
0·0	+1·000	+0·323	+0·191	+0·127	+0·091	+0·068	+0·053	+0·042	+0·034	+0·028	+0·024	+0·020	+0·017	+0·015	+0·013	+0·011	+0·010	+0·009	+0·008	+0·008	+0·007
0·5	+0·275	0·196	0·125	0·085	0·061	0·046	0·035	0·028	0·023	0·019	0·016	0·014	0·012	0·010	0·009	0·008	0·007	0·006	0·006	0·005	0·005
1·0	−0·076	+0·013	+0·029	+0·027	+0·023	+0·019	+0·015	+0·012	+0·010	+0·009	+0·008	+0·006	+0·006	+0·005	+0·004	+0·004	+0·003	+0·003	0·003	0·003	0·003
1·5	0·229	−0·111	−0·055	−0·028	−0·016	−0·009	−0·006	−0·004	0·003	−0·002	−0·001	0·001	−0·001	−0·001	−0·001	−0·000	−0·000	0·000	+0·000	+0·000	+0·000
2·0	0·280	0·176	0·111	0·072	0·049	0·035	0·026	0·019	0·015	0·012	0·010	0·008	0·007	0·006	0·005	0·005	0·004	0·003	0·003	−0·002	0·002
2·5	−0·281	−0·200	−0·142	−0·101	−0·074	−0·055	−0·042	−0·033	−0·027	−0·022	−0·018	−0·015	−0·013	−0·011	−0·010	0·008	−0·008	−0·007	−0·006	−0·005	−0·005
3·0	0·261	0·202	0·154	0·117	0·090	0·070	0·055	0·044	0·036	0·030	0·025	0·021	0·018	0·016	0·014	0·012	0·011	0·009	0·008	0·008	0·007
3·5	0·234	0·192	0·154	0·123	0·099	0·080	0·065	0·053	0·044	0·037	0·031	0·027	0·023	0·020	0·018	0·016	0·014	0·012	0·011	0·010	0·009
4·0	0·205	0·176	0·148	0·123	0·102	0·085	0·071	0·059	0·050	0·043	0·037	0·032	0·028	0·024	0·021	0·019	0·017	0·015	0·014	0·012	0·011
4·5	0·180	0·160	0·139	0·120	0·102	0·087	0·074	0·063	0·054	0·047	0·041	0·036	0·031	0·028	0·025	0·022	0·020	0·018	0·016	0·015	0·013
5·0	−0·157	−0·144	−0·129	−0·114	−0·099	−0·087	−0·075	−0·065	−0·057	−0·050	−0·044	−0·039	0·034	−0·030	−0·027	0·025	−0·022	−0·020	−0·018	0·017	−0·015
5·5	0·138	0·129	0·118	0·107	0·095	0·085	0·075	0·066	0·058	0·052	0·046	0·041	0·036	0·033	0·029	0·027	0·024	0·022	0·020	0·018	0·017
6·0	0·122	0·116	0·108	0·100	0·090	0·082	0·073	0·066	0·059	0·053	0·047	0·042	0·038	0·034	0·032	0·029	0·026	0·024	0·022	0·020	0·018
6·5	0·109	0·105	0·099	0·093	0·085	0·078	0·071	0·065	0·059	0·053	0·048	0·043	0·039	0·036	0·033	0·030	0·027	0·025	0·023	0·021	0·020
7·0	0·098	0·095	0·091	0·086	0·080	0·075	0·069	0·063	0·058	0·053	0·048	0·045	0·041	0·037	0·034	0·031	0·029	0·026	0·024	0·022	0·021
7·5	−0·089	−0·087	−0·084	−0·080	−0·076	−0·071	−0·066	−0·061	−0·056	−0·052	−0·048	−0·044	−0·041	−0·038	−0·035	−0·032	−0·029	−0·027	−0·025	−0·023	−0·022
8·0	0·082	0·080	0·076	0·075	0·071	0·067	0·063	0·059	0·055	0·051	0·048	0·044	0·041	0·038	0·035	0·033	0·030	0·028	0·026	0·024	0·023
8·5	0·075	0·074	0·072	0·070	0·067	0·064	0·061	0·057	0·054	0·050	0·047	0·044	0·041	0·038	0·035	0·033	0·031	0·028	0·027	0·025	0·023
9·0	0·070	0·068	0·067	0·066	0·063	0·061	0·058	0·055	0·052	0·049	0·046	0·043	0·040	0·037	0·035	0·033	0·031	0·029	0·027	0·025	0·024
9·5	0·064	0·064	0·063	0·062	0·060	0·058	0·055	0·053	0·050	0·047	0·045	0·042	0·040	0·037	0·035	0·033	0·031	0·029	0·027	0·026	0·024
10·0	−0·060	−0·061	−0·060	−0·058	−0·056	−0·055	−0·053	−0·051	−0·048	−0·046	−0·043	−0·041	−0·040	−0·037	−0·035	−0·033	−0·031	−0·029	−0·027	−0·026	−0·024

Table 1B

Imaginary Part of $F(\omega)$

$-\mathscr{I}(\omega)\rightarrow$

$\mathscr{R}(\omega)\downarrow$	0·0	0·5	1·0	1·5	2·0	2·5	3·0	3·5	4·0	4·5	5·0	5·5	6·0	6·5	7·0	7·5	8·0	8·5	9·0	9·5	10·0
0·0	−0·000	−0·268	−0·232	−0·198	−0·171	−0·150	0·132	−0·119	−0·107	0·098	0·089	0·083	0·077	0·072	−0·067	0·063	−0·060	0·057	−0·054	−0·051	−0·048
0·5	0·760	0·436	0·306	0·238	0·195	0·165	0·143	0·126	0·113	0·102	0·093	0·085	0·079	0·073	0·069	0·064	0·061	0·057	0·054	0·051	0·049
1·0	0·652	0·454	0·330	0·255	0·207	0·174	0·149	0·131	0·116	0·105	0·095	0·087	0·080	0·074	0·069	0·065	0·061	0·058	0·054	0·052	0·049
1·5	0·484	0·390	0·308	0·248	0·205	0·173	0·149	0·131	0·117	0·105	0·095	0·087	0·080	0·075	0·070	0·065	0·061	0·058	0·055	0·052	0·049
2·0	0·339	0·306	0·263	0·223	0·191	0·165	0·144	0·128	0·114	0·103	0·094	0·086	0·080	0·074	0·069	0·065	0·061	0·058	0·055	0·052	0·049
2·5	−0·230	−0·230	−0·213	−0·192	−0·170	−0·151	−0·135	−0·121	−0·110	−0·100	−0·092	−0·084	−0·078	−0·073	−0·068	−0·064	−0·060	−0·057	−0·054	−0·051	−0·049
3·0	0·153	0·168	0·167	0·159	0·147	0·135	0·123	0·113	0·103	0·095	0·088	0·081	0·076	0·071	0·067	0·063	0·059	0·057	0·054	0·051	0·048
3·5	0·100	0·121	0·129	0·129	0·125	0·118	0·110	0·103	0·096	0·089	0·083	0·078	0·073	0·068	0·065	0·061	0·058	0·055	0·053	0·050	0·048
4·0	0·065	0·087	0·099	0·104	0·104	0·102	0·097	0·093	0·088	0·083	0·078	0·073	0·069	0·066	0·062	0·059	0·056	0·054	0·051	0·049	0·047
4·5	0·042	0·062	0·075	0·083	0·086	0·087	0·085	0·083	0·079	0·076	0·072	0·069	0·066	0·062	0·059	0·057	0·055	0·052	0·050	0·048	0·046
5·0	−0·027	−0·045	−0·058	−0·066	−0·071	−0·074	−0·074	−0·073	−0·071	−0·069	−0·067	−0·064	−0·062	−0·059	−0·057	−0·055	−0·052	−0·050	−0·048	−0·046	−0·044
5·5	0·017	0·033	0·044	0·053	0·059	0·062	0·064	0·064	0·064	0·063	0·061	0·059	0·057	0·055	0·054	0·052	0·050	0·048	0·046	0·045	0·043
6·0	0·011	0·024	0·035	0·043	0·049	0·053	0·055	0·057	0·057	0·057	0·056	0·055	0·053	0·052	0·051	0·049	0·048	0·046	0·044	0·043	0·042
6·5	0·007	0·018	0·027	0·035	0·041	0·045	0·048	0·050	0·051	0·051	0·051	0·050	0·050	0·049	0·047	0·046	0·045	0·044	0·042	0·041	0·040
7·0	0·004	0·014	0·022	0·029	0·034	0·039	0·042	0·044	0·045	0·046	0·046	0·046	0·046	0·045	0·044	0·044	0·043	0·041	0·040	0·039	0·038
7·5	−0·003	−0·011	−0·018	−0·024	−0·029	−0·033	−0·036	−0·039	−0·040	−0·042	−0·042	−0·042	−0·042	−0·042	−0·041	−0·041	−0·040	−0·039	−0·039	−0·038	−0·037
8·0	0·002	0·008	0·015	0·020	0·025	0·029	0·032	0·034	0·036	0·037	0·038	0·039	0·039	0·039	0·039	0·038	0·038	0·037	0·036	0·036	0·035
8·5	0·001	0·007	0·012	0·017	0·021	0·025	0·028	0·030	0·032	0·034	0·035	0·035	0·036	0·036	0·036	0·036	0·036	0·035	0·035	0·034	0·034
9·0	0·001	0·005	0·010	0·014	0·018	0·021	0·024	0·027	0·029	0·030	0·032	0·033	0·033	0·033	0·034	0·034	0·034	0·033	0·033	0·033	0·032
9·5	0·000	0·004	0·008	0·013	0·016	0·019	0·022	0·024	0·026	0·028	0·029	0·030	0·031	0·031	0·031	0·032	0·032	0·031	0·031	0·031	0·031
10·0	+0·000	−0·004	−0·008	−0·011	−0·013	−0·017	−0·019	−0·022	−0·023	−0·025	−0·025	−0·027	−0·028	−0·029	−0·029	−0·030	−0·030	−0·030	−0·030	−0·029	−0·029

Table 2A

Real Part of $F(\omega)$

$-\mathscr{I}(\omega)\rightarrow$

$\mathscr{R}(\omega)\downarrow$	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0	1·1	1·2	1·3	1·4	1·5	1·6	1·7	1·8	1·9	2·0
0·0	+1·000	+0·632	+0·509	+0·428	+0·369	+0·323	+0·286	+0·256	+0·231	+0·209	+0·191	+0·174	+0·160	+0·148	+0·137	+0·127	+0·118	+0·110	+0·103	+0·096	+0·091
0·1	0·813	0·627	0·503	0·420	0·359	0·313	0·276	0·246	0·221	0·200	0·182	0·166	0·152	0·140	0·130	0·120	0·112	0·104	0·097	0·091	0·086
0·2	0·649	0·549	0·460	0·390	0·336	0·294	0·260	0·231	0·208	0·188	0·171	0·156	0·143	0·132	0·122	0·113	0·105	0·098	0·091	0·085	0·080
0·3	0·507	0·453	0·396	0·345	0·302	0·266	0·237	0·212	0·191	0·173	0·157	0·144	0·132	0·122	0·112	0·104	0·097	0·090	0·084	0·079	0·074
0·4	0·383	0·357	0·325	0·291	0·261	0·233	0·210	0·189	0·171	0·156	0·142	0·130	0·120	0·110	0·102	0·095	0·088	0·082	0·077	0·072	0·068
0·5	+0·275	+0·269	+0·254	+0·235	+0·215	+0·196	+0·179	+0·163	+0·149	+0·136	+0·125	+0·115	+0·106	+0·098	+0·091	+0·085	+0·079	+0·074	+0·069	+0·065	+0·061
0·6	0·182	0·190	0·188	0·179	0·170	0·158	0·146	0·135	0·125	0·115	0·107	0·099	0·092	0·085	0·079	0·074	0·069	0·065	0·061	0·057	0·054
0·7	0·102	0·119	0·126	0·127	0·124	0·119	0·113	0·107	0·100	0·094	0·088	0·082	0·076	0·071	0·067	0·063	0·059	0·055	0·052	0·049	0·046
0·8	+0·033	0·055	0·059	0·077	0·081	0·082	0·081	0·078	0·075	0·072	0·068	0·064	0·061	0·057	0·054	0·051	0·048	0·046	0·043	0·041	0·039
0·9	−0·026	+0·000	+0·020	+0·033	0·041	0·047	0·049	0·051	0·051	0·050	0·049	0·047	0·045	0·043	0·041	0·039	0·037	0·036	0·034	0·032	0·031
1·0	−0·078	−0·047	−0·025	0·008	+0·005	+0·013	+0·020	+0·024	+0·027	+0·028	+0·029	+0·030	+0·029	+0·029	+0·028	+0·027	+0·027	+0·026	+0·025	+0·024	+0·023
1·1	0·122	0·089	0·064	0·044	−0·029	−0·017	−0·008	−0·001	+0·004	+0·008	+0·011	+0·013	+0·014	0·015	0·015	0·016	0·016	0·016	0·016	0·015	0·015
1·2	0·157	0·124	0·097	0·076	0·058	0·045	0·033	0·024	−0·017	−0·012	0·007	−0·004	−0·001	+0·001	+0·003	+0·004	+0·005	+0·006	+0·006	+0·007	+0·007
1·3	0·185	0·153	0·126	0·104	0·085	0·069	0·056	0·046	0·037	0·030	0·024	0·019	0·015	−0·012	−0·009	0·007	−0·005	−0·004	−0·003	−0·002	−0·001
1·4	0·209	0·178	0·151	0·128	0·108	0·091	0·077	0·065	0·055	0·047	0·040	0·034	0·029	0·025	0·021	0·018	0·015	0·013	0·011	0·010	0·008
1·5	−0·229	0·199	−0·172	0·149	−0·129	0·111	0·096	−0·083	−0·072	0·063	−0·055	−0·048	−0·042	−0·037	−0·032	−0·028	−0·025	−0·022	−0·020	−0·018	−0·016
1·6	0·245	0·216	0·189	0·166	0·146	0·128	0·113	0·099	0·087	0·077	0·068	0·060	0·054	0·048	0·043	0·038	0·034	0·031	0·028	0·025	0·023
1·7	0·258	0·229	0·204	0·181	0·161	0·143	0·127	0·113	0·101	0·090	0·081	0·072	0·065	0·058	0·053	0·048	0·043	0·039	0·036	0·033	0·030
1·8	0·268	0·241	0·216	0·194	0·174	0·156	0·140	0·126	0·113	0·102	0·092	0·083	0·075	0·068	0·062	0·056	0·052	0·047	0·043	0·040	0·037
1·9	0·275	0·249	0·225	0·204	0·184	0·167	0·151	0·136	0·124	0·112	0·102	0·093	0·084	0·077	0·070	0·065	0·059	0·055	0·050	0·046	0·043
2·0	−0·280	−0·255	−0·233	−0·212	−0·193	−0·176	−0·160	−0·146	−0·133	−0·121	−0·111	−0·101	−0·093	−0·085	−0·078	−0·072	−0·067	−0·061	−0·057	−0·053	−0·049

Table 2B

Imaginary Part of $F(\omega)$

$-\mathscr{I}(\omega)\rightarrow$

$\mathscr{R}(\omega)\downarrow$	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0	1·1	1·2	1·3	1·4	1·5	1·6	1·7	1·8	1·9	2·0
0·0	−0·000	−0·234	−0·265	−0·273	−0·273	−0·268	−0·262	−0·255	−0·248	−0·240	−0·232	−0·225	−0·218	−0·211	−0·204	−0·198	−0·192	−0·187	−0·181	−0·176	−0·171
0·1	0·507	0·403	0·365	0·343	0·325	0·309	0·295	0·282	0·271	0·259	0·249	0·240	0·231	0·222	0·215	0·207	0·200	0·194	0·188	0·182	0·177
0·2	0·649	0·528	0·456	0·409	0·375	0·349	0·327	0·309	0·293	0·278	0·266	0·254	0·243	0·233	0·224	0·216	0·208	0·201	0·194	0·188	0·182
0·3	0·719	0·605	0·522	0·463	0·419	0·385	0·357	0·333	0·314	0·296	0·281	0·267	0·255	0·244	0·233	0·224	0·215	0·207	0·200	0·193	0·187
0·4	0·751	0·648	0·566	0·503	0·453	0·414	0·381	0·354	0·332	0·312	0·294	0·279	0·265	0·253	0·242	0·231	0·222	0·213	0·205	0·198	0·191
0·5	−0·760	−0·668	−0·591	−0·529	−0·478	−0·436	−0·401	−0·371	−0·347	−0·325	−0·306	−0·289	−0·274	−0·261	−0·249	−0·238	−0·228	−0·219	−0·210	−0·202	−0·195
0·6	0·753	0·672	0·602	0·543	0·492	0·450	0·415	0·384	0·358	0·335	0·315	0·298	0·282	0·268	0·255	0·244	0·233	0·223	0·214	0·206	0·199
0·7	0·736	0·666	0·602	0·548	0·500	0·459	0·424	0·393	0·367	0·343	0·322	0·304	0·288	0·273	0·260	0·248	0·237	0·227	0·218	0·209	0·201
0·8	0·712	0·651	0·595	0·545	0·501	0·462	0·428	0·398	0·372	0·348	0·327	0·309	0·292	0·277	0·264	0·252	0·240	0·230	0·221	0·212	0·204
0·9	0·684	0·631	0·582	0·537	0·496	0·460	0·428	0·399	0·373	0·350	0·330	0·312	0·295	0·280	0·266	0·254	0·243	0·232	0·223	0·214	0·206
1·0	0·652	−0·607	−0·564	−0·524	−0·488	−0·454	−0·424	−0·397	−0·373	−0·350	−0·330	−0·312	−0·296	−0·281	−0·268	−0·255	−0·244	−0·234	−0·224	−0·215	−0·207
1·1	0·619	0·580	0·543	0·508	0·476	0·445	0·418	0·392	0·369	0·348	0·329	0·312	0·296	0·281	0·268	0·256	0·244	0·234	0·225	0·216	0·208
1·2	0·585	0·552	0·520	0·490	0·461	0·434	0·409	0·385	0·364	0·344	0·326	0·309	0·294	0·280	0·267	0·255	0·244	0·234	0·224	0·216	0·208
1·3	0·551	0·524	0·496	0·470	0·445	0·420	0·398	0·376	0·357	0·338	0·321	0·305	0·291	0·277	0·265	0·253	0·243	0·233	0·224	0·215	0·207
1·4	0·517	0·495	0·472	0·449	0·427	0·406	0·385	0·366	0·348	0·331	0·315	0·300	0·287	0·274	0·262	0·251	0·241	0·231	0·222	0·214	0·206
1·5	−0·484	−0·466	−0·447	−0·428	−0·408	−0·390	−0·372	−0·355	−0·338	−0·323	−0·308	−0·295	−0·282	−0·270	−0·258	−0·248	−0·238	−0·229	−0·220	−0·212	−0·205
1·6	0·453	0·438	0·422	0·406	0·389	0·373	0·358	0·342	0·328	0·314	0·300	0·288	0·276	0·265	0·254	0·244	0·235	0·226	0·218	0·210	0·203
1·7	0·422	0·410	0·398	0·384	0·370	0·356	0·343	0·329	0·316	0·304	0·292	0·280	0·269	0·259	0·249	0·240	0·231	0·222	0·215	0·207	0·200
1·8	0·393	0·384	0·374	0·363	0·351	0·340	0·328	0·316	0·305	0·293	0·283	0·272	0·262	0·252	0·243	0·235	0·226	0·219	0·211	0·204	0·197
1·9	0·365	0·359	0·351	0·342	0·333	0·323	0·313	0·303	0·293	0·283	0·273	0·264	0·254	0·246	0·237	0·229	0·222	0·214	0·207	0·201	0·194
2·0	−0·339	−0·335	−0·329	−0·322	−0·314	−0·306	−0·298	−0·289	−0·280	−0·272	−0·263	−0·255	−0·247	−0·239	−0·231	−0·223	−0·216	−0·210	−0·203	−0·197	−0·191

Table 3A

REAL PART OF $F(\omega)$

$\mathscr{R}(\omega)\downarrow$ / $-\mathscr{I}(\omega)\rightarrow$	0·00	0·05	0·10	0·15	0·20	0·25	0·30	0·35	0·40	0·45	0·50
0·00	+1·000	+0·731	+0·632	+0·563	+0·509	+0·465	−0·428	+0·397	+0·369	+0·345	+0·323
0·05	0·903	0·750	0·644	0·569	0·511	0·465	0·427	0·394	0·366	0·341	0·319
0·10	0·813	0·713	0·627	0·558	0·503	0·458	0·420	0·388	0·359	0·335	0·313
0·15	0·728	0·658	0·592	0·535	0·485	0·443	0·407	0·377	0·350	0·326	0·305
0·20	0·649	0·599	0·549	0·502	0·460	0·423	0·390	0·362	0·336	0·314	0·294
0·25	+0·576	+0·539	+0·501	+0·464	+0·429	+0·398	+0·369	+0·343	+0·320	+0·300	+0·281
0·30	0·507	0·481	0·453	0·424	0·396	0·370	0·345	0·323	0·302	0·283	0·266
0·35	0·443	0·425	0·404	0·383	0·361	0·339	0·319	0·300	0·282	0·266	0·250
0·40	0·383	0·372	0·357	0·342	0·325	0·308	0·291	0·276	0·261	0·246	0·233
0·45	0·327	0·321	0·312	0·301	0·289	0·276	0·263	0·251	0·238	0·226	0·215
0·50	+0·275	+0·274	+0·269	+0·262	+0·254	+0·245	+0·235	+0·225	+0·215	+0·206	+0·196

Table 3B

IMAGINARY PART OF $F(\omega)$

$\mathscr{R}(\omega)\downarrow$ / $-\mathscr{I}(\omega)\rightarrow$	0·00	0·05	0·10	0·15	0·20	0·25	0·30	0·35	0·40	0·45	0·50
0·00	−0·000	−0·194	−0·234	−0·254	−0·265	−0·271	−0·273	−0·273	−0·273	−0·271	−0·268
0·05	0·377	0·329	0·320	0·317	0·315	0·312	0·308	0·303	0·299	0·294	0·289
0·10	0·507	0·440	0·403	0·381	0·365	0·353	0·343	0·334	0·325	0·317	0·309
0·15	0·591	0·521	0·473	0·438	0·413	0·393	0·377	0·363	0·351	0·340	0·330
0·20	0·649	0·581	0·528	0·487	0·456	0·430	0·409	0·391	0·375	0·362	0·349
0·25	−0·690	−0·625	−0·571	−0·528	−0·492	−0·463	−0·438	−0·417	−0·398	−0·382	−0·368
0·30	0·719	0·657	0·605	0·560	0·522	0·490	0·463	0·440	0·419	0·401	0·385
0·35	0·739	0·681	0·630	0·585	0·547	0·514	0·485	0·460	0·437	0·418	0·400
0·40	0·751	0·697	0·648	0·604	0·566	0·532	0·503	0·477	0·453	0·432	0·414
0·45	0·758	0·707	0·660	0·618	0·581	0·547	0·517	0·491	0·467	0·445	0·426
0·50	−0·760	−0·712	−0·668	−0·628	−0·591	−0·558	−0·529	−0·502	−0·478	−0·456	−0·436

REFERENCES

) SOMMERFELD, A.: "On the Spreading of Waves in Wireless Telegraphy," *Annalen der Physik*, 1909, **28**, p. 665.

) NORTON, K. A.: "The Propagation of Radio Waves over the Surface of the Earth and in the Upper Atmosphere: Part I," *Proceedings of the Institute of Radio Engineers*, 1936, **24**, p. 1367.

) MCPETRIE, J. S., and STICKLAND, A. C.: "Reflection Curves and Propagation Characteristics of Radio Waves along the Earth's Surface," *Journal I.E.E.*, 1940, **87**, p. 135.

) NORTON, K. A.: "The Calculation of Ground Wave Field Intensity over a Finitely Conducting Spherical Earth," *Proceedings of the Institute of Radio Engineers*, 1941, **29**, p. 623.

(5) BURROWS, C. R.: "Radio Propagation over Plane Earth—Field Strength Curves," *Bell System Technical Journal*, 1937, **16**, p. 45.

(6) MCPETRIE, J. S.: "The Reflection Coefficient of the Earth's Surface for Radio Waves," *Journal I.E.E.*, 1938, **82**, p. 214.

(7) CLEMMOW, P. C., and MUMFORD, CARA M.: "A Table of $\sqrt{(\frac{1}{2}\pi)}e^{\frac{1}{2}i\pi\rho^2}\int_{\rho}^{\infty} e^{-\frac{1}{2}i\pi\lambda^2}d\lambda$ for Complex Values of ρ," *Philosophical Transactions of the Royal Society of London, A*, 1952, **245**, p. 189.

621.318.1.042.2 : 621.3.013.5

Monograph No. 1
Jan. 19

LEAKAGE FLUX AND SURFACE POLARITY IN IRON RING STAMPINGS

By P. HAMMOND, M.A., Associate Member.

(*The paper was first received 6th July, and in revised form 29th September*, 1954. *It was published as an* INSTITUTION MONOGRAPH *in January*, 1955.)

SUMMARY

The paper deals with the conduction of magnetic flux through iron circuits magnetized by windings which surround only a part of the iron.

The complete solution is obtained for an infinitely long hollow iron cylinder magnetized by a single current loop, and it is shown that surface polarity on the iron plays a dominant part in causing the flux to be substantially constant at all sections around the circumference of the cylinder.

Experimental evidence is brought forward in support of the theoretical investigation.

(1) INTRODUCTION

If an annular ring is completely and uniformly wound with a magnetizing winding, the magnetic force at a given radius inside the ring is constant; accordingly, there is the same total flux through every cross-section of the ring. Moreover, there is no external field, if one neglects the second-order effect due to the fact that the winding is a helix and not a true current sheet. This simple distribution of flux makes the annular ring an ideal form for carrying out tests to measure the magnetic qualities of the material.

Although it is natural to think of a ring made of iron, there has so far been no need to specify the material of the ring. However, if the magnetizing winding is now bunched together, so that the ring is magnetized by a concentrated coil instead of a uniformly distributed winding, experiment shows that the behaviour of iron is completely different from that of non-magnetic material. This difference is shown up most clearly by the fact that the flux around the ring is substantially constant when an iron ring is used, but shows no sign of constancy when the ring is non-magnetic. The paper discusses this phenomenon from the viewpoint of magnetic theory and puts forward a theoretical analysis backed by experimental results.

(2) THEORETICAL INVESTIGATION

(2.1) Qualitative Analysis

It is very surprising that the flux distribution around an annular iron ring is substantially constant regardless of the disposition of the magnetizing winding. The problem raised by this constancy is often dismissed by the statement that the iron conducts flux because of its high permeability. This is undoubtedly correct, but scarcely deserves to rank as an explanation, if by explanation is meant that anyone who knew that iron has a high permeability should have foreseen the result. The treatment in the paper is based on the discussion of the problem given by Moullin,* who shows that the surface polarity on the iron provides the mechanism by which the iron can maintain a substantially constant flux. Moullin dealt with the case of an iron ring and pointed out that it would be extremely difficult mathematically to analyse the flux distribution for such a ring, even if it were assumed that the material had constant permeability. It was therefore decided to examine the proble of a long tube of material of constant permeability, for which mathematical solution was readily obtainable. It then becar possible to check experimentally whether such a solution cou give any help in solving the problem of an iron tube whc permeability was not constant.

Before proceeding with the mathematical analysis it may helpful to examine the problem in general terms by means of numerical example.

Consider an iron ring of mean diameter 10cm and let it made from round bar of 1cm radius. If the ring is wou with a uniformly distributed magnetizing winding of 100 A the magnetic force is equal to 4 oersteds anywhere on the me circumference of the ring. If the magnetizing winding is th bunched together into a closely wound coil, the magnetic for in the absence of the iron will be 20π oersteds at the centre the coil and $\pi/100$ oersteds at the mean diameter of the ring a point diametrically opposite to the concentrated coil. Fro Fig. 1 it will be seen that the magnetic force due to the ma

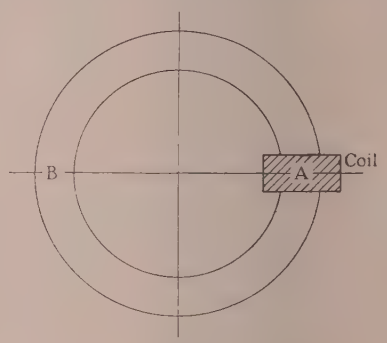


Fig. 1

metizing current is 2 000 times as strong at A as at B. Neverth less, if the ring is made of iron, the flux density is substantial the same at A as at B, and, moreover, it has the value appropria to a uniform magnetizing winding. Hence the iron has reduc the magnetic force at A from 20π to 4 oersteds and increas that at B from $\pi/100$ to 4 oersteds. How has this come abou

It is sometimes argued that this problem is fundamental the same as that of current conduction from a battery throu a wire. But this analogy is not a good one, because in curre conduction the electric force is transmitted by an actual mov ment of charge, whereas there is certainly no movement magnetic matter through iron. Moreover, in conductors the is a volume as well as a surface distribution of charge, but the is no such volume distribution of magnetism so long as t permeability is constant.

Another analogy that is often put forward is that the ir can be regarded as a chain of compass needles placed end end. Such a picture is not in accord with the domain theory ferromagnetism. It must be stressed that if there is to be certain flux density in iron, there must also be a certain magne force acting at that point. If permanent magnetism (hysteres

* MOULLIN, E. B.: "Principles of Electromagnetism" (University Press, Oxford), Third edition, pp. 164–168.

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Hammond is in the Department of Engineering, University of Cambridge.

neglected, the only way in which this magnetic force can arise by the appearance of surface polarity on the iron.

It follows that the change in magnetic force produced by the n ring must be due to surface polarity on the iron and that is surface polarity must act even when the permeability of the n becomes infinitely large. If there is to be surface polarity ere must also be flux emerging from the iron, and the field can longer be zero outside the iron, as was the case with the uniformly wound ring. Flux that leaves the iron is commonly called eakage" flux, and in many applications this leakage flux is garded as an imperfection in the design of the apparatus. ur discussion, however, has shown that the leakage flux is a anifestation of the mechanism of surface polarity by which e iron maintains a constant flux around the ring. (We shall ll this the mutual flux, because it links both the cross-sections and B in Fig. 1.) It would be quite accurate to state that thout the leakage flux there would be hardly any mutual flux.

If the permeability of the iron is very large, a small surface larity will be sufficient to maintain a substantially constant ux around the iron. The ratio of leakage flux to mutual flux ill therefore decrease with increase in permeability and will nd to zero when the permeability tends to infinity, but it is ear that the absolute value of the leakage flux will not nd to zero. Consider the action of the surface polarity with ference to Fig. 2, in which the leakage-flux lines are shown

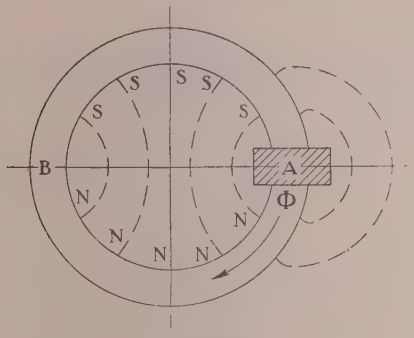


Fig. 2

iagrammatically. The surface polarity is north polar over the alf-circumference ANB and south polar over the half-circumrence ASB; the general direction of the magnetic force is from ie lower half to the upper half. It can now be seen that the eduction in magnetic force at A is due to the fact that the surface olarity opposes the magnetic force due to the magnetizing coil. t B, on the other hand, the surface polarity aids the magnetic orce due to the magnetizing coil and so the magnetic force at B increased. Thus the surface polarity has the required tendency f making the magnetic force more uniform around the ring ian it would otherwise have been.

(2.2) Quantitative Analysis

In order to arrive at a simple mathematical solution of the roblem which can be tested experimentally, it is desirable to npose two simplifying restrictions. The first is that the problem ould be treated as a 2-dimensional one. This means that the nchor ring must be replaced by an infinitely long cylinder. econdly, the permeability of the material must be taken as onstant and the solution must therefore be regarded as only a eneral guide to the behaviour of iron cylinders. But there is ason to expect that the discrepancy will not be very serious, nd this is borne out by the experimental results. The solution is indeed of considerable practical value, because it enables us to form an estimate of the magnitude of the leakage flux for iron at different permeabilities. This leakage flux will always be small and may well be of no consequence in transformers intended for use in pointer instruments. But transformers are required in certain types of calculating machine, and in such applications it may be desirable to estimate the order of the discrepancy between voltage ratio and turns ratio.

(2.2.1) Magnetization by Single Conductor Inside the Cylinder.

Fig. 3 shows a cylindrical tube of internal radius b and external radius a. A current I flows in a round rod of radius

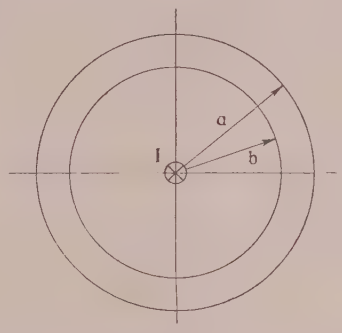


Fig. 3

smaller than b coaxial with the tube. Then H is constant at any radius around the iron tube, its magnitude varying between $2I/b$ and $2I/a$ between the inside and outside surfaces. The field outside the iron is the same as it would be if the iron were not there. It would not be proper to call this external flux a leakage flux, since it does not represent a leakage of flux out of the iron into the air. There is no surface polarity on the iron cylinder. The conductor carrying the return current is considered to be so far away that its field is insignificant. Such a condition can readily be achieved in practice. Consider now the altered condition illustrated in Fig. 4, in which the current is

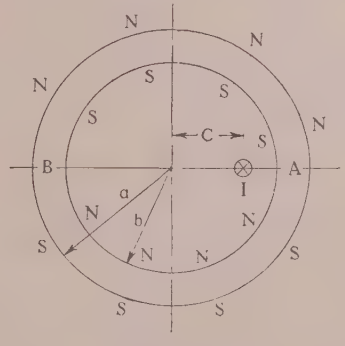


Fig. 4

no longer coaxial with the tube but is displaced from the centre through a distance c. It is shown in Section 6.1 that the flux passing the cross-section A is related to the flux at B by the expression

$$\frac{\Phi_A}{\Phi_B} = \frac{\mu \log \frac{a}{b} - 2 \log \left(1 - \frac{c}{b}\right)}{\mu \log \frac{a}{b} - 2 \log \left(1 + \frac{c}{b}\right)} = k \text{ (say)} \quad . \quad . \quad (1)$$

so long as $\mu \gg 1$—a condition fulfilled by irons of widely differing magnetic quality. This equation enables the leakage flux to be calculated from a knowledge of the dimensions of the cylinder, the permeability of the iron and the eccentricity of the conductor. For small eccentricities

$$k \simeq 1 + \frac{4\frac{c}{b}}{\mu \log \frac{a}{b}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Thus the fractional leakage flux is directly proportional to the eccentricity and inversely proportional to the permeability. Another case of interest is the one in which the conductor lies close against the iron. If, for example, $c/b = 95/100$ and $a/b = 1{\cdot}5$, substitution in eqn. (1) shows that

$$k = 1 + \frac{4{\cdot}95}{\mu} \log \frac{195}{5} = 1 + \frac{18{\cdot}1}{\mu} \quad . \quad . \quad . \quad (3)$$

Hence the leakage flux in such a case will exceed 1% unless μ exceeds about 2 000.

In Section 6.1 it is shown that the surface polarity is related to the radial magnetic force by the expression

$$2\pi\sigma = \frac{\mu - 1}{\mu + 1} H_r$$

It is thus clear that for high permeabilities the surface polarity will rapidly tend to a constant finite value and will not disappear even when the permeability becomes infinite. This is a perfectly general statement about all surface polarity.

Consider now the internal surface polarity (see Fig. 4). It has been derived in Section 6.1 and is given by

$$\sigma_b = -\frac{Iy}{\pi} \sum_1^\infty \frac{1 - yx^{2n}}{1 - y^2x^{2n}} \frac{c^n}{b^{n+1}} \sin n\theta \quad . \quad . \quad . \quad (4)$$

where $y = \dfrac{\mu - 1}{\mu + 1}$ and $x = \dfrac{b}{a}$

If $\mu \gg 1, \quad y \to 1$

and

$$\sigma_b = -\frac{I}{\pi} \sum_1^\infty \frac{c^n}{b^{n+1}} \sin n\theta \quad . \quad . \quad . \quad . \quad (5)$$

$= -1/2\pi \times$ (radial component of field due to the current I at radius b in the absence of iron). The potential function of this surface polarity is given by

$$V = -2I \sum_1^\infty \frac{c^n}{nr^n} \sin n\theta \quad . \quad . \quad . \quad . \quad . \quad (6)$$

for $\mu \gg 1$ and the potential function of the magnetic field due to the current I for $r > c$ is

$$V = 2I\left(\theta + \sum_1^\infty \frac{c^n}{nr^n} \sin n\theta\right) \quad . \quad . \quad . \quad . \quad (7)$$

It is therefore seen that the leakage flux due to the internal surface polarity is finite, even if the permeability becomes infinitely large, and that it tends rapidly to a constant value which is independent of the external radius a. Consideration of the potential functions reveals that, for high permeabilities, the internal surface polarity entirely cancels the radial magnetic force of the current and modifies the tangential magnetic force so that it has the value it would have if the current were placed at the origin of co-ordinates and the iron tube were removed. This condition is that of perfect magnetic screening, and it is shown below that the external surface polarity is zero for such a case, as would be expected. Consideration of the circuital la $\oint H.dl = 4\pi I$ shows that the tangential magnetic force can nev be entirely "screened off," although the radial magnetic force reduced to zero.

The surface polarity at radius a, i.e. on the outer surface the tube, is derived in Section 6.1 and is shown to be

$$\sigma_a = \frac{2Iy}{\pi(\mu + 1)} \sum_1^\infty \frac{1}{1 - y^2x^{2n}} \frac{c^n}{a^{n+1}} \sin n\theta \quad . \quad . \quad (1$$

if $\mu \gg 1, \quad y \to 1$

and

$$\sigma_a \simeq \frac{2I}{\pi\mu} \sum_1^\infty \frac{1}{1 - x^{2n}} \frac{c^n}{a^{n+1}} \sin n\theta \quad . \quad . \quad . \quad (1$$

Thus, since $x < 1$,

$$\sigma_a < \frac{2I}{\pi\mu} \sum_1^\infty \frac{c^n}{a^{n+1}} \sin n\theta \quad . \quad . \quad . \quad . \quad (1$$

Therefore σ_a is less than the product of $1/\pi\mu$ and the radi component of field due to current I at radius a.

It is also seen that the polarity σ_a will tend to zero as μ increased. This is as expected and should be compared wi the constant value of σ_b at large values of μ. We have show therefore, that the leakage flux is predominantly caused by t internal surface polarity.

(2.2.2) **Magnetization by Single Conductor Outside the Iron Tube.**

The surface polarity for the case of magnetization shown Fig. 5 is derived in Section 6.2 and is shown to be σ_b on the insi of the tube, where

$$\sigma_b = \frac{2Iy}{\pi(\mu + 1)} \sum_1^\infty \frac{1}{1 - y^2x^{2n}} \frac{b^{n-1}}{d^n} \sin n\theta \quad . \quad . \quad (1$$

If $\mu \gg 1, \quad y \to 1$

and

$$\sigma_b = \frac{2I}{\pi\mu} \sum_1^\infty \frac{1}{1 - x^{2n}} \frac{b^{n-1}}{d^n} \sin n\theta \quad . \quad . \quad . \quad (1$$

and hence

$$\sigma_b < \frac{2I}{\pi\mu} \sum_1^\infty \frac{b^{n-1}}{d^n} \sin n\theta \quad . \quad . \quad . \quad . \quad (1$$

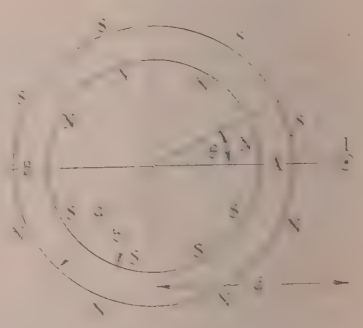


Fig. 5

i.e. σ_b is less than the product of $1/\pi\mu$ and the radial compone of field at radius b in the absence of iron. It is therefore see that the internal polarity will tend to zero as μ is increase The external polarity is given by

$$\sigma_a = \frac{Iy}{\pi} \sum_1^\infty \frac{1 - yx^{2n}}{1 - y^2x^{2n}} \frac{a^{n-1}}{d^n} \sin n\theta \quad . \quad . \quad . \quad (1$$

nd if $\mu \gg 1$

$$\sigma_a = \frac{I}{\pi} \sum_1^\infty \frac{a^{n-1}}{d^n} \sin n\theta \quad . \quad . \quad . \quad . \quad (15)$$

$= 1/2\pi \times$ (radial component of field at radius a in the bsence of iron).

The external polarity therefore has a constant finite value, ven when the permeability becomes infinite. Consideration of ne potential functions reveals that under this condition the xternal polarity completely cancels the field of the current or any radius $r < a$. This condition is thus one of perfect agnetic screening, and it is of interest that there will be neither adial nor tangential field inside the tube. It was pointed out n Section 2.2.1 that the tangential field of an internal current annot be screened off by surrounding the current with an iron ube.

Because there exists external surface polarity on the tube even hen the permeability tends to infinity, there will be "leakage" ux emerging from the iron. The extreme variation of "mutual" ux crossing the wall of the tube is derived in Section 6.2 and given by:

$$\frac{\Phi_A}{\Phi_B} = \frac{\log 1 - \frac{a}{d}}{\log 1 + \frac{a}{d}} \quad . \quad . \quad . \quad . \quad . \quad (16)$$

hen $\mu \gg 1$.

It is rather surprising that, not only the fractional leakage ux, but the mutual flux itself is independent of both the permeability and the thickness of the tube. The flux carried by the all of the tube is, in fact, exactly twice the flux that would have rossed the radius a if the tube had been removed. It might be bjected that the absence of magnetic force inside the tube i.e. $r < a$) should also lead to an absence of flux carried by the ube. This is a very sound objection and is entirely in accord ith the discussion of Section 2.1. The reason for this apparent ontradiction is that by letting μ tend to infinity we are really o longer treating this as a problem in magnetism. Our discussion would apply rigorously to the problem of a perfectly onducting cylinder in an electric field. Such a cylinder could e treated as the limiting case of a dielectric cylinder when the ermittivity tended to infinity. But there is no process in magetism analogous to conduction in electricity, and the idea of nagnetic flux density without magnetic force must be rejected.

So long as our minds are clear on this matter, however, it is seful to examine the limiting case of infinite permeability, which lthough physically impossible, simplifies the mathematics and ives the mutual and leakage fluxes to a high degree of pproximation.

Consider now a conductor close to the outside wall of the tube. f $a/d = 95/100$

$$\frac{\Phi_A}{\Phi_B} = \frac{\log \frac{5}{100}}{\log \frac{195}{100}} = -4{\cdot}5$$

The reason for the minus sign is clear when it is considered that he flux will be directed in opposite senses at A and B. It is, in act, undesirable to speak here of "mutual" or "leakage" flux, ecause the leakage flux is sufficiently large to reverse the mutual lux at opposite ends of a diameter of the tube. It should be ept in mind that these fluxes which are associated with magetization by an external current are considerably less than hose associated with an internal current. The ratio between hem is of the order of the permeability.

(2.2.3) Tube magnetized by Equal and Opposite Currents, One Inside and One Outside.

It is now desirable to superpose the results of the last two Sections in order to obtain the case of an iron tube magnetized by a single loop of current.

Let the equal and opposite currents be situated as shown in Fig. 6. The internal current causes more flux to cross section A

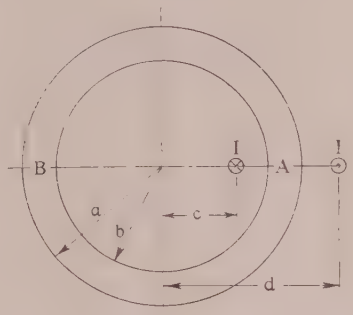


Fig. 6

than section B. The external current causes at A a flux directed in the same sense as the flux there due to the internal current. At B, however, the flux due to the external current opposes the flux due to the internal current. It is therefore clear that the provision of the "return conductor" outside the tube accentuates the variation of flux around the tube; in other words, it increases the leakage flux.

Combining the results obtained in Section 6.1 and 6.2 we have

$$\frac{\Phi}{2I} = \mu\mu_0 \log \frac{a}{b} + 2\mu_0 \sum_1^\infty \frac{c^n}{nb^n} \cos n\theta + 2\mu_0 \sum_1^\infty \frac{a^n}{nd^n} \cos n\theta \quad . \quad (17)$$

Therefore

$$\frac{\Phi_A}{\Phi_B} = \frac{\mu \log \frac{a}{b} - 2 \log \left(1 - \frac{c}{b}\right) - 2 \log \left(1 - \frac{a}{d}\right)}{\mu \log \frac{a}{b} - 2 \log \left(1 + \frac{c}{b}\right) - 2 \log \left(1 + \frac{a}{d}\right)} \quad . \quad (18)$$

If there is only slight eccentricity of the internal conductor and if the return conductor is at a great distance from the tube, we have $c/b \ll 1$ and $a/d \ll 1$.

Therefore

$$\frac{\Phi_A}{\Phi_B} \simeq 1 + \frac{4\left(\frac{c}{b} + \frac{a}{d}\right)}{\mu \log \frac{a}{b}} \quad . \quad . \quad . \quad . \quad (19)$$

The case of greatest interest arises when both wires are almost at the surfaces of the tube.

If we take $c/b = a/d = 95/100$

then

$$\frac{\Phi_A}{\Phi_B} = \frac{\mu \log \frac{a}{b} + 12}{\mu \log \frac{a}{b} - 2{\cdot}68}$$

and if $a/b = 3/2$

$$\frac{\Phi_A}{\Phi_B} = \frac{\mu + 29{\cdot}7}{\mu - 6{\cdot}62} \simeq 1 + \frac{36{\cdot}3}{\mu}$$

Hence the leakage flux will exceed 1% unless μ exceeds 3 600. Fig. 7 shows a family of curves derived on the basis of $a/b = 3/2$ for varying eccentricities ($c/b = a/d$).

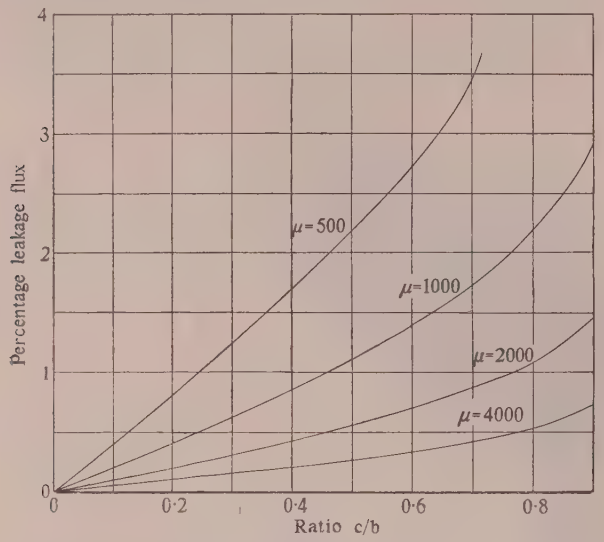


Fig. 7.—Theoretical leakage flux.

(2.2.4) Examination of the Tangential Magnetic Force Inside the Tube magnetized by a Single Loop of Current.

It has already been pointed out that it is not helpful to dismiss the problem of flux distribution in iron by the statement that iron is a good conductor of flux, since, in fact, this so-called process of conduction is radically different from that of conduction of electric currents. We have stressed that for any flux density B to arise anywhere in space, whether inside a material or outside, it is necessary that there should be the requisite magnetic force H at that point. This magnetic force can be calculated by the use of the inverse-square law between point magnetic poles, and this law is independent of the medium in which or through which the magnetic force acts. The behaviour of iron as a good conductor of flux, even when it is magnetized by an unsymmetrical magnetizing winding, is therefore entirely attributable to its polarizability, i.e. to the fact that pole strength will be induced on its surface when it is placed in a magnetic field.

So far we have directed attention to this surface polarity and to the flux distribution around the tube. This flux distribution is of great interest, because it provides a ready means of examining experimentally whether the theoretical results apply to actual iron tubes. But in this Section it is proposed to examine the variation of magnetic force around the tube, so as to gain a closer insight into the mechanism by which the flux density remains substantially constant around the tube, even when this is magnetized by a single loop of current.

At any radius r in the wall of the tube ($b < r < a$) the tangential force due to the internal current and the polarities induced by it is given in Section 6.1 as

$$-\frac{H_\theta}{2I} = \frac{1}{r} + \frac{2}{\mu+1}\sum_1^\infty \frac{1}{1-y^2x^{2n}}\left(\frac{1}{r^{n+1}} + \frac{yr^{n-1}}{a^{2n}}\right)c^n \cos n\theta \quad . \quad (20)$$

Similarly, the tangential force due to the external current, and the polarities induced by it, are given in Section 6.2 as

$$-\frac{H_\theta}{2I} = \frac{2}{\mu+1}\sum_1^\infty \frac{1}{1-y^2x^{2n}}\left(r^{n-1} + y\frac{b^{2n}}{r^{n+1}}\right)\frac{1}{d^n}\cos n\theta \quad . \quad (21)$$

Hence the total tangential force is given by

$$-\frac{H_\theta}{2I} = \frac{1}{r} + \frac{2}{\mu+1}\sum_1^\infty \frac{1}{1-y^2x^{2n}}\left(\frac{c^n}{r^{n+1}} + \frac{yb^{2n}}{d^n r^{n+1}} + \frac{yr^{n-1}c^n}{a^{2n}} + \frac{r^{n-1}}{d^n}\right)\cos n\theta \quad . \quad (2$$

If $\mu \gg 1$

$$-\frac{H_\theta}{2I} = \frac{1}{r} + \frac{2}{\mu}\sum_1^\infty \frac{1}{1-x^{2n}}\left[\left(c^n + \frac{b^{2n}}{d^n}\right)\frac{1}{r^{n+1}} + \left(\frac{c^n}{a^{2n}} + \frac{1}{d^n}\right)r^{n-1}\right]\cos n\theta \quad . \quad (2$$

$$= \frac{1}{r} + \frac{2}{\mu}\sum_1^\infty \frac{1}{1-x^n}\left(\frac{c^n}{r^{n+1}} + \frac{r^{n-1}}{d^n}\right)\cos n\theta \quad . \quad . \quad . \quad (2$$

if $c/b = a/d$.

Now $-H_\theta/2I = 1/r$ would be the value of H_θ if the tube we magnetized by a uniform current sheet of strength I. Hen any variation of H_θ caused by the concentration of the ma netizing winding into a single coil is given by the term

$$\frac{2}{\mu}\sum_1^\infty \frac{1}{1-x^n}\left(\frac{c^n}{r^{n+1}} + \frac{r^{n-1}}{d^n}\right)\cos n\theta$$

and this is less than

$$\frac{2}{\mu}\sum_1^\infty \left(\frac{c^n}{r^{n+1}} + \frac{r^{n-1}}{d^n}\right)\cos n\theta$$

i.e. less than

$$\frac{2}{\mu} \times \text{radial component of field due to single current loop in the absence of the tube.}$$

Compare this now with the field of the single current loop the absence of the tube:

$$-\frac{H_\theta}{2I} = \frac{1}{r} + \sum_1^\infty \frac{c^n}{r^{n+1}}\cos n\theta + \sum_1^\infty \frac{r^{n-1}}{d^n}\cos n\theta \quad . \quad (2$$

This comparison shows clearly that the variation of tangenti magnetic force has been reduced in the ratio which is more th $1 : \mu/2$. This reduction must be due to the fact that the magne force arising from the surface polarity opposes that from t current where the latter is strong, because of its proximity the current loop, and aids it where it is weak.

The tangential magnetic force due to the induced polariti is given by

$$-\frac{H_\theta}{2I} = \sum_1^\infty \frac{1}{1-y^2x^{2n}}\left[\frac{2y}{\mu+1}\frac{c^n r^{n-1}}{a^{2n}} - y(1-yx^{2n})\frac{c^n}{r^{n+1}} - y(1-yx^{2n})\frac{r^{n-1}}{d^n} + \frac{2y}{\mu+1}\frac{b^{2n}}{d^n r^{n+1}}\right]\cos n\theta \quad . \quad (2$$

If $c/b = a/d$ and $\mu \gg 1$,

$$-\frac{H_\theta}{2I} = \sum_1^\infty \frac{1}{1-x^{2n}}\left[\frac{r^{n-1}}{d^n}\left(\frac{2}{\mu}x^n - 1 + x^{2n}\right) - \frac{c^n}{r^{n+1}}\left(1 - x^{2n} - \frac{2}{\mu}x^n\right)\right]\cos n\theta$$

$$= \sum_1^\infty \left[\frac{r^{n-1}}{d^n}\left(\frac{2}{\mu}\frac{x^n}{1-x^{2n}} - 1\right) + \frac{c^n}{r^{n+1}}\left(\frac{2}{\mu}\frac{x^n}{1-x^{2n}} - 1\right)\right]\cos n\theta \quad . \quad . \quad . \quad (2$$

ence the action of the surface polarity is to reduce H_θ at $\theta = 0$ om

$$-\frac{H_\theta}{2I} = \frac{1}{r} + \sum_1^\infty \left(\frac{c^n}{r^{n+1}} + \frac{r^{n-1}}{d^n}\right) \quad . \quad . \quad . \quad (28)$$

$$-\frac{H_\theta}{2I} = \frac{1}{r} + \frac{2}{\mu}\sum_1^\infty \frac{x^n}{1-x^{2n}}\left(\frac{c^n}{r^{n+1}} + \frac{r^{n-1}}{d^n}\right) \quad . \quad (29)$$

e. from $$-\frac{H_\theta}{2I} = \frac{1}{r} + \frac{c}{r(r-c)} + \frac{1}{d-r} \quad . \quad . \quad . \quad (30)$$

$$-\frac{H_\theta}{2I} = \frac{1}{r} + \frac{2}{\mu}\left[\frac{cx}{r(r-cx)} + \frac{x}{d-rx}\right] \quad . \quad . \quad (31)$$

o long as $x^{2n} \to 0$.

At $\theta = \pi$ the tangential force is increased from

$$-\frac{H_\theta}{2I} = \frac{1}{r} - \frac{c}{r(r+c)} - \frac{1}{d+r} \quad . \quad . \quad . \quad (32)$$

$$-\frac{H_\theta}{2I} = \frac{1}{r} - \frac{2}{\mu}\frac{cx}{r(r+cx)} + \frac{x}{d+rx} \quad . \quad . \quad (33)$$

f we consider H_θ at the mean radius $r = (c+d)/2$ and if $= b/a = 2/3$ and $c/b = a/d = 0{\cdot}95$, we have an extreme ariation of $H_\theta/2I$, in the absence of the tube, from

$$-\frac{H_\theta}{2I} = \frac{8}{r} \text{ to } -\frac{H_\theta}{2I} = \frac{0{\cdot}127}{r}$$

When the tube is placed in position this variation is reduced and he relevant values are

$$-\frac{H_\theta}{2I} = \frac{1}{r}\left(1 + \frac{4{\cdot}28}{\mu}\right) \text{ and } -\frac{H_\theta}{2I} = \frac{1}{r}\left(1 - \frac{1{\cdot}36}{\mu}\right)$$

f $\mu = 2\,000$ the effect of the surface polarity on the tube is to educe the range of variation of H_θ from approximately 1 : 63 o approximately 1 : 1·00282. Thus it comes about that the lux density is substantially constant around the tube.

(3) EXPERIMENTAL INVESTIGATION

(3.1) Apparatus

The iron tubes were built up from silicon-iron stampings, ightly insulated on one side, the outside and inside diameters eing 11·75 and 9·2cm respectively. Three different lengths of ube were used, in order to obtain some indication of the effect f flux leakage at the ends, the approximate lengths being 35, 17·5 and 8·75cm. The tubes were mounted with their axes orizontal on a slide which could be moved transversely to a ong copper rod of diameter 0·475cm, which carried the 50c/s magnetizing current. The eccentricity of the conductor was aried by moving the iron tube. Two search coils of 44-gauge .s.c. copper wire were wound on each of the tubes at opposite nds of a diameter, and the ends of these coils were brought ut to terminals so that the coils could be connected in series or n opposition. The numbers of turns on the search coils were 50, 100 and 200 respectively for the three tubes. The output from these coils was connected to an integrating circuit consisting of a 39μF condenser in series with a 4 920-ohm resistor. The condenser terminals were connected to a cathode-ray oscillograph.

(3.2) Method of Measurement

The "return" part of the magnetizing winding was sufficiently far from the iron tubes not to affect the measurements, and the magnetization could thus be attributed entirely to the long straight conductor. External inductance was inserted into the magnetizing circuit so as to keep the magnetizing current close to a pure sine wave. Fig. 8(*a*) shows a typical current trace obtained on the cathode-ray oscillograph for a current of 100amp.

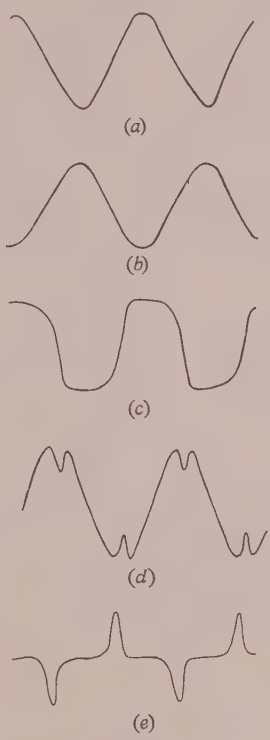


Fig. 8.—Waveforms.
(*a*) Current.
(*b*) Leakage flux.
(*c*) Mutual flux.
(*d*) "Leakage" voltage.
(*e*) "Mutual" voltage.

In order to obtain the leakage-flux curve, the search coils were connected in opposition, and Fig. 8(*b*) shows a typical trace. Fig. 8(*c*) shows the mutual-flux curve, obtained with the search coils in series. Figs. 8(*d*) and 8(*e*) were obtained by connecting the search coils straight to the oscillograph, and omitting the integrating circuit. They show, therefore, the "leakage" and the "mutual" voltage waves respectively.

In order to try to eliminate the end-leakage effects of the tubes it was decided to use "guard rings," i.e. to pack the ends with additional stampings. Tests were made on the two shorter tubes packed like this at one end only and at both ends. A considerable reduction in leakage flux was observed, but difficulties were encountered because the leakage flux was exceedingly sensitive to slight differences in the packing of the ends. Owing to the bulge of the search coils this packing could not be made perfect, and considerably varying readings could be obtained by altering the axial pressure to which the tubes were subjected.

In addition to the a.c. measurements, two d.c. magnetic reversal curves were obtained by the use of a fluxmeter. One curve was obtained in the usual way by winding a specimen of approximately square cross-section with a uniformly distributed magnetizing winding and search coil. The other curve was obtained by using the 8·75cm tube with its search coils and reversing a direct current in the long copper conductor. These

two methods gave approximately identical results, and any small differences were no doubt due to slight variations in the magnetic characteristics of the different stampings.

(3.3) Discussion of Results

It has been shown in Section 2.2 that the surface polarity, and therefore the leakage flux, is approximately independent of μ so long as $\mu \gg 1$. This is borne out by Fig. 8(*b*), where it is seen that the leakage-flux wave has exactly the shape of the magnetizing-current wave. The effect of supplying a magnetizing-current wave containing harmonics was tried, and in every case the leakage-flux wave contained the same harmonics. Figs. 8(*c*) and 8(*e*) show that there was considerable saturation in the iron. It is, of course, to be expected that the leakage flux will show hardly any signs of saturation, even when the iron is very highly saturated, because the path of the leakage flux is predominantly in air. However, it is not until one considers the surface polarity that the mechanism by which this is achieved becomes at all clear.

Since the leakage-flux wave is sinusoidal it was possible to calibrate the oscillograph to a scale of flux density, and this made it possible to obtain the maximum flux density of the total flux wave. Hence the fractional leakage flux could be obtained for various eccentricities of the magnetizing conductor. The maximum total flux density afforded a further check on the d.c. reversal curve and reasonable agreement was found.

Fig. 9 shows a typical family of curves of fractional leakage flux obtained with the 35cm tube at different magnetizing currents and eccentricities of the conductor carrying this current. In order to give a ready comparison with the theoretical curves shown in Fig. 7, the theoretical curve has been included on Fig. 9; it is plotted as $100\mu(k-1)$, where

$$\frac{\Phi_A}{\Phi_B} = \frac{\mu \log \frac{a}{b} - 2 \log \left(1 - \frac{c}{b}\right)}{\mu \log \frac{a}{b} - 2 \log \left(1 + \frac{c}{b}\right)} = k \text{ (say)} \quad . \quad . \quad (1)$$

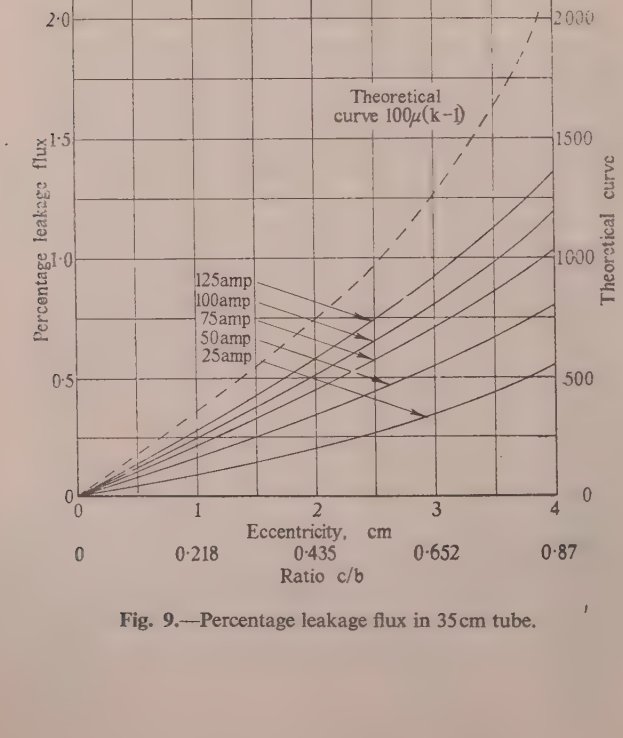


Fig. 9.—Percentage leakage flux in 35cm tube.

It is clear from Fig. 9 that the theoretical curve is of the same shape as the experimental curves. In order to test the closeness of fit, Fig. 10 has been plotted. A curve at low saturation

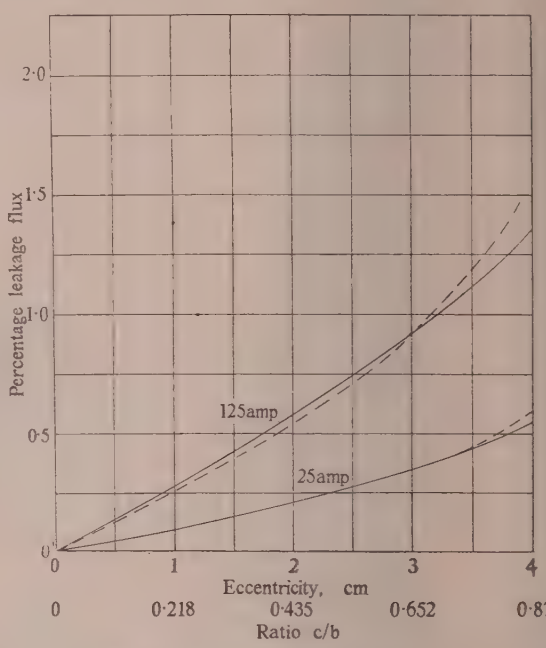


Fig. 10.—Comparison of shape of theoretical and experimental leakage-flux curves.

(25amp magnetizing current) and another at high saturation (125amp magnetizing current) have been chosen. The broken curves are the theoretical ones, which have been derived by assuming a constant permeability such that the experimental and theoretical curves coincide at an eccentricity of 3cm. At such an eccentricity $100\mu(k-1) = 815 \log 1{\cdot}652/0{\cdot}348 = 1\,270$. This gives an "equivalent μ" of 3 640 for the 25amp curve and 1 380 for the 125amp curve. It can be seen that the chief difference between the theoretical and experimental curves is the greater steepness of the former. This is not surprising when it is realized that the theoretical curve is calculated for a tube of infinite length. The end effect is bound to increase the leakage of the relatively short tubes used in the experiments, and is likely to be most marked when the iron is highly saturated and when the eccentricity of the magnetizing conductor is great.

Results similar to those shown in Fig. 9 were obtained with the tubes 17·5 and 8·75cm long, both when these tubes were used by themselves and when they were packed with additional stampings as described in Section 3.2. Equivalent permeabilities were again obtained by equating the theoretical and experimental leakage fluxes at an eccentricity of 3cm. These equivalent permeabilities are plotted against magnetic force in Fig. 11, and on this diagram a similar curve of μ against H was plotted as obtained from a d.c. magnetic reversal curve. Fig. 11 shows how the leakage flux is considerably reduced by increasing the length of the tube and by packing the ends.

Table 1 gives the ratio between the equivalent permeability and the permeability determined from d.c. test for the 8·75 and 17·5cm tubes when these are packed with additional stamping at both ends.

The equivalent permeability is about 20% less than the actual permeability, and the theoretical treatment is thus shown to be substantially applicable to iron tubes. A part at least of the 20%

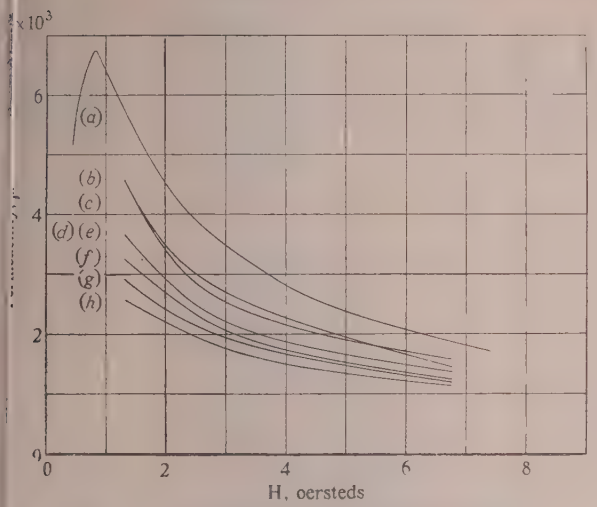


Fig. 11.—Comparison of equivalent permeabilities.

(*a*) D.C. reversal curve.
(*b*) 8·75cm specimen packed both sides.
(*c*) 17·5cm specimen packed both sides.
(*d*) 35cm specimen.
(*e*) 17·5cm specimen packed one side.
(*f*) 8·75cm specimen packed one side.
(*g*) 17·5cm specimen.
(*h*) 8·75cm specimen.

Table 1

H	Equivalent μ/Actual μ	
	8·75cm tube	17·5cm tube
oersteds		
1·35	0·80	0·79
2·7	0·76	0·71
4·05	0·80	0·74
5·4	0·81	0·84
6·75	0·77	0·80

iscrepancy can be accounted for when it is understood that here will be some end leakage, even when the tubes have their nds packed, because this packing cannot be made perfect see Section 3.2). A further slight reduction in the equivalent ermeability will be due to eddy currents, and there will also be ome effect from the permanent magnetism (hysteresis) of the ron. The theoretical treatment neglects hysteresis.

A test was carried out on the 35cm tube when this was magetized by an external conductor as discussed in Section 2.2.2. he theoretical flux ratio is

$$\frac{\Phi_A}{\Phi_B} = \frac{\log 1 - \frac{a}{d}}{\log 1 + \frac{a}{d}} \quad . \quad . \quad . \quad . \quad . \quad (16)$$

n this case a/d was 0·96 when the conductor was touching the ube. Hence $\Phi_A/\Phi_B = -4{\cdot}8$ theoretically, and the average xperimental value was found to be −4·3. This was largely ndependent of saturation.

(4) CONCLUSION

It has been shown that surface polarity (or leakage flux) gives o iron its remarkable property of being able to maintain a ubstantially constant total flux around an iron circuit which is magnetized by a concentrated coil. Good experimental agreement is obtained with the theoretical treatment, even when the apparently drastic assumption of constant permeability in the iron has been made. It therefore becomes possible to estimate the magnitude of the leakage flux in any particular set of iron ring stampings. Furthermore, it becomes possible to estimate the order of the flux leakage for more complicated iron circuits.

(5) ACKNOWLEDGMENT

The author gratefully acknowledges the guidance and encouragement given to him by Professor E. B. Moullin.

(6) APPENDIX

(6.1) *To calculate the surface distribution of polarity induced on a tube of material of constant permeability by a current flowing parallel to the axis of the tube and at a distance* c *from the axis, where* c *is smaller than the internal radius of the tube.*

We require first the field of the current in the absence of the tube. This is given by the potential function

$$V = 2I\left(\theta + \sum_1^\infty \frac{c^n}{nr^n} \sin n\theta\right) \quad . \quad . \quad . \quad . \quad (7)$$

$$[r > c]^*$$

Let the induced surface polarity be denoted by σ_a on the outside surface of the tube and σ_b on the inside surface. Consideration will show that σ_a will be positive (north polarity) over the top half of the tube and negative (south polarity) over the bottom half; these conditions will be reversed for σ_b. Hence σ_a and σ_b can be expressed by the Fourier series

$$\sigma_a = \sum_1^\infty \sigma_{an} \sin n\theta \quad . \quad . \quad . \quad . \quad . \quad (34)$$

and

$$\sigma_b = -\sum_1^\infty \sigma_{bn} \sin n\theta \quad . \quad . \quad . \quad . \quad . \quad (35)$$

Also the potential function of σ_a for $r < a$ will be

$$V = \sum_1^\infty \frac{2\pi r^n}{na^{n-1}} \sigma_{an} \sin n\theta \quad . \quad . \quad . \quad . \quad (36)$$

and the potential function of σ_b for $r > b$ will be

$$V = -\sum_1^\infty \frac{2\pi b^{n+1}}{nr^n} \sigma_{bn} \sin n\theta \quad . \quad . \quad . \quad . \quad (37)$$

Consider now the radial field at the surface of a cylindrical shell of polarity.

Let the circle shown in Fig. 12 represent the cross-section of such a cylindrical shell, which carries a completely general distribution of polarity $\sigma = f(\theta)$. It is required to find the force at a typical point P on the circumference of the shell due to this polarity. Now a filament situated at Q produces at P a force in

* In order to shorten the paper, no detailed derivation of the potential functions has been given. It will be clear that these potentials must satisfy Laplace's equation, which is merely a mathematical statement following from a consideration of the inverse-square law. Laplace's equation in cylindrical co-ordinates is

$$\frac{\partial^2 V}{\partial r^2} + \frac{1}{r}\frac{\partial V}{\partial r} + \frac{1}{r^2}\frac{\partial^2 V}{\partial \theta^2} = 0$$

and the solution will therefore be of the form

$$V = r^n(A \cos n\theta + B \sin n\theta) + \frac{1}{r^n}(C \cos n\theta + D \sin n\theta).$$

All this work is based on Clerk Maxwell, and the particular cases discussed in the paper are fully dealt with by Moullin:

Clerk Maxwell, J.: "Electricity and Magnetism" (Clarendon Press, Oxford), Part IV, Chapter 1.

Moullin, E. B.: "Principles of Electromagnetism" (University Press, Oxford), Third edition, pp. 197 and 403.

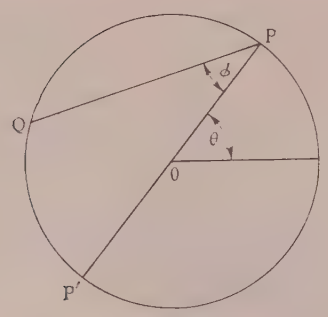


Fig. 12.—Shell of polarity.

the direction QP of magnitude $\delta H = 2R\delta\theta f(\theta)/QP$. The radial component of this force is

$$\delta H_r = \frac{2R\delta\theta f(\theta)}{QP}\cos\phi = \frac{2R\delta\theta f(\theta)}{QP}\frac{QP}{PP'} = \delta\theta f(\theta) \quad . \quad (38)$$

Hence the contribution of radial force made by the filament at Q is independent of the position of Q and depends only on the magnitude of the polarity at that point. Hence the total radial force at P is given by

$$H_r = \int_0^{2\pi} f(\theta)d\theta = \int_0^{2\pi} \sigma d\theta \quad . \quad . \quad . \quad . \quad (39)$$

If the polarity is induced there will be equal amounts of positive and negative polarity on the shell, and hence the radial force at any point in the shell will be zero. At a point very close to the shell and just outside it, the radial force will depend on the local polarity and will be the same as for an infinite plane, i.e. $2\pi\sigma$. Just inside the shell the radial force will be $-2\pi\sigma$. Having established this property of cylinders, we can determine the surface polarity on the cylindrical tube. By continuity of radial force at radius a

$$H_r + 2\pi\sigma_a = \mu(H_r - 2\pi\sigma_a)$$

Therefore

$$2\pi\sigma_a = \frac{\mu - 1}{\mu + 1}H_r = yH_r \text{ (say)} \quad . \quad . \quad . \quad (40)$$

where H_r is the radial force at $r = a$ due to the current I and the surface polarity σ_b.

Hence

$$H_r = -\frac{\partial}{\partial r}\left[2I\left(\theta + \sum_1^\infty \frac{c^n}{nr^n}\sin n\theta\right) - \frac{2\pi b^{n+1}}{nr^n}\sigma_{bn}\sin n\theta\right]$$

at $r = a$. . . . . (41)

Therefore $$\frac{2\pi\sigma_{an}}{y} = 2I\frac{c^n}{a^{n+1}} - \frac{2\pi b^{n+1}}{a^{n+1}}\sigma_{bn} \quad . \quad . \quad . \quad . \quad (42)$$

If $b/a = x$ (say)

$$2\pi\sigma_{an} = 2Iy\frac{c^n}{a^{n+1}} - 2\pi yx^{n+1}\sigma_{bn} \quad . \quad . \quad . \quad (43)$$

Similarly $$2\pi\sigma_{bn} = 2Iy\frac{c^n}{b^{n+1}} - 2\pi yx^{n-1}\sigma_{an} \quad . \quad . \quad . \quad (44)$$

Hence $$\sigma_{an} = \frac{2Iy}{\pi(\mu + 1)}\frac{1}{1 - y^2x^{2n}}\frac{c^n}{a^{n+1}} \quad . \quad . \quad . \quad (45)$$

and $$\sigma_{bn} = \frac{Iy}{\pi}\frac{1 - yx^{2n}}{1 - y^2x^{2n}}\frac{c^n}{b^{n+1}} \quad . \quad . \quad . \quad . \quad (46)$$

(6.2) *To calculate the fractional leakage flux between t cross-sections A and B of the tube, as indicated in Fig. 4.*

The tangential magnetic force in the tube is due to the current and the induced surface polarities σ_a and σ_b. Its value can derived from the potential functions by the relationship

$$H_\theta = -\frac{1}{r}\frac{\partial V}{\partial\theta}$$

Therefore

$$H_\theta = -\frac{2I}{r}\left(1 + \sum_1^\infty \frac{c^n}{r^n}\cos n\theta\right) - \frac{1}{r}\sum_1^\infty \frac{2\pi r^n}{a^{n-1}}\sigma_{an}\cos n\theta + \frac{1}{r}\sum_1^\infty \frac{2\pi b^{n+1}}{r^n}\sigma_{bn}\cos n\theta \quad . \quad (4$$

Therefore

$$-\frac{H_\theta}{2I} = \frac{1}{r} + \sum_1^\infty \cos n\theta\left(\frac{c^n}{r^{n+1}} + \frac{2y}{\mu + 1}\frac{c^n r^{n-1}}{a^{2n}}\frac{1}{1 - y^2x^{2n}} - y\frac{c^n}{r^{n-1}}\frac{1 - yx^{2n}}{1 - y^2x^{2n}}\right) = \frac{1}{r} + \frac{2}{\mu + 1}\sum_1^\infty \frac{1}{1 - y^2x^{2n}}\left(\frac{1}{r^{n+1}} + \frac{yr^{n-1}}{a^{2n}}\right)c^n\cos n\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2$$

Thus

$$\frac{\Phi}{2I} = \frac{-\int_b^a \mu\mu_0 H_\theta dr}{2I} = \mu\mu_0\log\frac{a}{b} + \frac{2\mu\mu_0}{\mu + 1}\sum_1^\infty \frac{1 - x^n}{1 - yx^n}\frac{c^n}{nb^n}\cos$$

. . . . . (4

If $\mu \gg 1$

$$\frac{\Phi}{2I} = \mu\mu_0\log\frac{a}{b} + 2\mu_0\sum_1^\infty \frac{c^n}{nb^n}\cos n\theta$$

$$= \mu\mu_0\log\frac{a}{b} - \mu_0\log\left(1 - 2\frac{c}{b}\cos\theta + \frac{c^2}{b^2}\right) \quad . \quad (4$$

Hence $$\frac{\Phi_A}{\Phi_B} = \frac{\mu\log\frac{a}{b} - 2\log\left(1 - \frac{c}{b}\right)}{\mu\log\frac{a}{b} - 2\log\left(1 + \frac{c}{b}\right)} \quad . \quad . \quad . \quad ($$

(6.3) *To calculate the surface distribution of polarity induc on a tube of material of constant permeability μ by a curre flowing parallel to the axis of the tube and at a distance c from t axis, where c is larger than the external radius of the tube.*

Let the external and internal induced polarities be denoted before by

$$\sigma_a = \sum_1^\infty \sigma_{an}\sin n\theta \quad . \quad . \quad . \quad . \quad . \quad (3$$

and $$\sigma_b = -\sum_1^\infty \sigma_{bn}\sin n\theta \quad . \quad . \quad . \quad . \quad . \quad (3$$

hence their potential functions will be

$$V = \sum_1^\infty \frac{2\pi r^n}{na^{n-1}}\sigma_{an}\sin n\theta \quad . \quad . \quad . \quad . \quad (3$$

when $r < a$ and

$$V = -\sum_1^\infty \frac{2\pi b^{n+1}}{nr^n}\sigma_{bn}\sin n\theta \quad . \quad . \quad . \quad . \quad (3$$

when $r > b$.

o the potential function of the current I will be

$$V = -2I \sum_1^\infty \frac{r^n}{nd^n} \sin n\theta \quad . \quad . \quad . \quad . \quad (50)$$

ere $r < d$.

continuity of radial force at radius a and b,

$$2\pi\sigma_{an} = 2Iy\frac{a^{n-1}}{d^n} - 2\pi y x^{n+1}\sigma_{bn} \quad . \quad . \quad . \quad (51)$$

$$2\pi\sigma_{bn} = 2Iy\frac{b^{n-1}}{d^n} - 2\pi y x^{n-1}\sigma_{an} \quad . \quad . \quad . \quad (52)$$

nce $$\sigma_a = \frac{Iy}{\pi} \sum_1^\infty \frac{1 - yx^{2n}}{1 - y^2x^{2n}} \frac{a^{n-1}}{d^n} \sin n\theta \quad . \quad . \quad . \quad (14)$$

$$\sigma_b = \frac{2Iy}{\pi(\mu + 1)} \sum_1^\infty \frac{1}{1 - y^2x^{2n}} \frac{b^{n-1}}{d^n} \sin n\theta \quad . \quad . \quad (11)$$

6.4) *To calculate the fractional leakage flux between the ss-sections A and B of the tube, as indicated in Fig. 5.*

before, $$H_\theta = -\frac{1}{r}\frac{\partial V}{\partial \theta}$$

erefore

$$= 2I \sum_1^\infty \frac{r^{n-1}}{d^n} \cos n\theta - \sum_1^\infty \frac{2\pi r^{n-1}}{a^{n-1}} \sigma_{an} \cos n\theta + \sum_1^\infty \frac{2\pi b^{n+1}}{r^{n+1}} \sigma_{bn} \cos n\theta \quad . \quad (53)$$

erefore

$$\frac{H_\theta}{2I} = \frac{2}{\mu + 1} \sum_1^\infty \frac{1}{d^n(1 - y^2x^{2n})} \left(r^{n-1} + y\frac{b^{2n}}{r^{n+1}}\right) \cos n\theta \quad . \quad (21)$$

Therefore

$$\frac{\Phi}{2I} = \frac{\int_b^a B_\theta dr}{2I} = \frac{2\mu\mu_0}{\mu + 1} \sum_1^\infty \frac{1}{nd^n(1 - y^2x^{2n})} \left[r^n - y\frac{b^{2n}}{r^n}\right]_b^a \cos n\theta \quad . \quad . \quad . \quad (54)$$

$$= \frac{2\mu\mu_0}{\mu + 1} \sum_1^\infty \frac{1 - yx^{2n} - \frac{2}{\mu + 1}x^n}{1 - y^2x^{2n}} \frac{a^n}{nd^n} \cos n\theta \quad . \quad . \quad . \quad (55)$$

$$\simeq 2\mu_0 \sum_1^\infty \frac{1 - x^{2n} - \frac{2}{\mu}x^n}{1 - x^{2n}} \frac{a^n}{nd^n} \cos n\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad (56)$$

when $\mu \gg 1$

$$\simeq 2\mu_0 \sum_1^\infty \frac{a^n}{nd^n} \cos n\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (57)$$

when $\mu \to \infty$

At $\theta = 0$, with this approximation,

$$\frac{\Phi}{2I} = 2\mu_0 \sum_1^\infty \frac{a^n}{nd^n} = 2\mu_0\left(\frac{a}{d} + \frac{a^2}{2d^2} + \ldots\right) \quad . \quad . \quad (58)$$

$$= -2\mu_0 \log\left(1 - \frac{a}{d}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (59)$$

and at $\theta = \pi$

$$\frac{\Phi}{2I} = -2\mu_0 \log\left(1 + \frac{a}{d}\right) \quad . \quad . \quad . \quad . \quad (60)$$

Hence $$\frac{\Phi_A}{\Phi_B} = \frac{\log 1 - \frac{a}{d}}{\log 1 + \frac{a}{d}} \quad . \quad . \quad . \quad . \quad . \quad (16)$$

DISCUSSION ON
"THE REFLECTION OF ELECTROMAGNETIC WAVES FROM A ROUGH SURFACE"*

Mr. Lee M. Spetner (*communicated*): I question the use of eqn. (27) to represent the conditional probability of finding a surface elevation of Z located at a distance r from a point where the surface height is known to be Z_0. Perhaps the author actually intended to assume a two-dimensional normal distribution for the two heights, Z and Z_0, in which case eqn. (27) should, instead, be

$$P(Z, Z_0, r) = \frac{1}{\sigma\sqrt{(2\pi)[1-\exp(-2r^2/a^2)]^{\frac{1}{2}}}} \exp\left\{-\frac{[Z - Z_0\exp(-r^2/a^2)]^2}{2\sigma^2[1-\exp(-2r^2/a^2)]}\right\}$$

This affects the expressions in eqns. (33) and (34) and eqn. (35) should become

$$\langle FF^*\rangle = \exp\left\{-\frac{2\pi i}{\lambda}[s(\sin\theta\cos\phi - \sin\psi) + t\sin\theta\sin\phi]\right\} \times \left\{\exp -\frac{4\pi^2\sigma^2}{\lambda^2}(\cos\theta + \cos\psi)^2[1-\exp(-r^2/a^2)] - \exp\left[-\frac{4\pi^2\sigma^2}{\lambda^2}(\cos\theta + \cos\psi)^2\right]\right\}$$

Using the approximation of eqn. (36), then, gives us for eqn. (37)

$$\langle FF^*\rangle = \exp\left\{-\frac{4\pi^2\sigma^2 r^2}{\lambda^2 a^2}(\cos\theta + \cos\psi)^2 - \frac{2\pi i}{\lambda}[s(\sin\theta\cos\phi - \sin\psi) + t\sin\theta\sin\phi)]\right\}$$

changing the exponent in the incoherent power by a factor of 2, giving for eqn. (38)

$$\langle|\mathscr{E} - \langle\mathscr{E}\rangle|^2\rangle = \frac{e^2}{4\lambda^2 r^2} lm \frac{\lambda^2 a^2}{4\pi\sigma^2} \exp\left[-\frac{a^2}{4\sigma^2}\frac{(\sin\theta\cos\phi - \sin\psi)^2 + \sin^2\theta\sin^2\phi}{(\cos\theta + \cos\psi)^2}\right]$$

where, incidentally, the term $(\cos\theta + \cos\psi)^2$ preceding the exponential in eqn. (38) actually is cancelled out in the integration over s and t.

Mr. H. Davies (*in reply*): The probability distribution of eqn. (27) is not satisfactory, as has been pointed out by Mr. Spetner, because it is not symmetrical in the two variables z and z_0. We can use a similar distribution which does have property, obtained by assuming a two-dimensional nor distribution including a correlation term. Keeping the s auto-correlation function, eqn. (27) will now read:

$$P(z, z_0, r) = \frac{1}{\sigma\sqrt{2\pi[1-\exp(-2r^2/a^2)]^{\frac{1}{2}}}} \exp\left\{-\frac{[z - z_0\exp(-r^2/a^2)}{2\sigma^2[1-\exp(-2r^2/a}\right.$$

the bivariate distribution now being

$$\phi(z, z_0, r) = \frac{1}{2\pi\sigma^2[1-\exp(-2r^2/a^2)]^{\frac{1}{2}}} \exp\left\{-\frac{z^2 + z_0^2 - 2zz_0\exp(-r^2/}{2\sigma^2[1-\exp(-2r^2/a^2)}\right.$$

This affects eqns. (33) and (34) slightly, and eqn. (35) sho now read

$$\langle FF^*\rangle = \exp\left\{-\frac{2\pi i}{\lambda}[s(\sin\theta\cos\phi - \sin\psi) + t\sin\theta\sin\phi]\right\} \times \left[\exp\left\{-\frac{4\pi^2\sigma^2}{\lambda^2}(\cos\theta + \cos\psi)^2[1-\exp(-r^2/a^2)]\right\} - \exp\left[-\frac{4\pi^2\sigma^2}{\lambda^2}(\cos\theta + \cos\psi)^2\right.\right.$$

Eqn. (37) is changed slightly and eqn. (38) becomes

$$\langle|\mathscr{E} - \langle\mathscr{E}\rangle|^2\rangle = \frac{e^2}{4\lambda^2 r^2} lm \frac{\lambda^2 a^2}{4\pi\sigma^2} \exp\left[-\frac{a^2}{4\sigma^2}\frac{(\sin\theta\cos\phi - \sin\psi)^2 + \sin^2\theta\sin^2}{(\cos\theta + \cos\psi)^2}\right.$$

and eqn. (39)

$$\langle|\mathscr{E} - \langle\mathscr{E}\rangle|^2\rangle = \frac{a^2 e^2 lm}{16\pi\sigma^2 r^2}\exp\left[-\frac{a^2}{16\sigma^2}(\chi^2 + w^2\sec^2\psi)\right.$$

There are similar slight changes in Section 4, giving for eqn.

$$P_s = \frac{P_t G A a^2 \tau c \beta}{128\pi^2\sigma^2 R_0^3}\operatorname{cosec}\psi\exp\left(-\frac{a^2}{4\sigma^2}\tan^2\psi\right)$$

Therefore the theoretical curves in Fig. 4 are the curves $a^2/2\sigma^2 = 50, 60$.

* DAVIES, H.: Monograph No. 90 R, January, 1954 (see **101**, Part IV, p. 209).

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

ISSUED IN THREE PARTS AS FOLLOWS:

Part A. POWER ENGINEERING (*February, April, etc.*)

Part B. RADIO AND ELECTRONIC ENGINEERING (*January, March, etc.*)

Part C. INSTITUTION MONOGRAPHS (*March and September only*)

PART C—MONOGRAPHS

MARCH 1955

CONTENTS OF THIS ISSUE

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part.
Example.—SMITH, J.: "Overhead Transmission Systems," *Proceedings I.E.E.*, Paper No. 3001 R, December, 1954 (102 A, p. 1234).

Published by The Institution, Savoy Place, London, W.C.2. *Telephone:* Temple Bar 7676. *Telegrams:* "Voltampere, Phone, London."
Printed by Unwin Brothers Limited, Woking and London.